SOME SIMPLE METHODS OF TESTING FOR FUNCTION FLUCTUATION

BY C. C. ANDERSON

Department of Education, University of Cambridge

I. *The concept of function fluctuation* (pp. 1–3). II. *Early methods of measuring function fluctuation* (pp. 3–6). III. *Thouless's correlation of differences* (pp. 6–8). IV. *Finney's analysis of variance model* (pp. 8–9). V. *Experimental study* (pp. 9–10). VI. *Conclusion* (p. 11). *References* (pp. 11–12).

I. THE CONCEPT OF FUNCTION FLUCTUATION

Thouless (1936) made three important contributions to the study of function fluctuation: first, he coined the term 'function fluctuation' itself; secondly, he was the first psychologist clearly to articulate its meaning in relation to test error; and thirdly, he devised methods by which the extent of its activity could be measured. He saw that lack of perfect self-consistency between parallel tests of the same mental function was determined partly by test error—the fact that the groups of items making up the two tests are not exactly measuring the same thing because of inevitable bias in the selection from the item population—and partly by a real variability in the function being measured, a tendency for which he proposed to reserve the term 'function fluctuation'.

> We may define the reliability of any measuring instrument as its degree of success in giving the same reading when applied to equal quantities. It may be tested by seeing how far the measuring instrument gives identical readings when applied twice to the same unvarying object....Absence of self-consistency in test results can, however, only be taken as an indication of unreliability as above defined if repetition of the test with the same individuals is really a measurement of the same unvarying mental function. This is not necessarily the case. Let us suppose that the ability, or whatever else we are attempting to measure, is in the same individual quantitatively different at different times. The fact that our test gives different results at different times is then no evidence that it is unreliable (that it gives different readings when used to measure quantities which are actually equal). Different test results would now follow even from the use of a perfectly reliable test, because the quantities measured are really different at different times of testing. (Thouless, 1936, pp. 325–6.)

The concept is not synonymous with crude variability in performance described in such statistics as Hull's (1927) variance of observed scores, or Hertzman's (1938) sum of all possible differences between scores or ranks. This would appear an elementary point, but recently Gulliksen (1950, p. 197) has confused the two because of an ambiguity in terminology. 'Methods of determining the instability of a trait as distinct from the instability of a test have been suggested by...Thouless (1936, 1939), Preston (1940), and Jackson & Ferguson (1941)'. However, the other authors cited are dealing with crude variability. Jackson & Ferguson are concerned entirely with obtaining a measure of test reliability in terms of analysis of variance instead of by the logically and statistically inferior method of correlations, and their gamma statistic is based on a comparison between the size of the errors of measurement and the size of the differences between the individuals tested. Preston attacks the situation in which subjects have taken a number of the same tests, or parallel tests of the same mental function, on different testing occasions. To obviate the labour of calculating for each individual the variance of his test scores, which might be taken as an index of what Preston calls 'trait variability', Preston (1940, p. 215) derives, in his own terms: 'an expression for the average variance of the distribution of individual scores from test to test, and an expression for the

variance of these variances in terms of the number of tests and the intercorrelations between them'. Gulliksen's confusion comes in part from Preston's belief that he is measuring trait variability, a term borrowed from Paulsen (1931) who uses it to mean function fluctuation, whereas Preston is in fact measuring crude variability. The two concepts are, of course, related, but are not identical.

Another semantic point, illustrating a possible ambiguity in Thouless's concept, has been raised by Finney (1953), who states that he prefers the term 'performance fluctuation', on grounds which he does not clearly specify. Of course it is obvious that what is hypothesized abstractly as mental functions can only be measured in terms of mental products. E. L. Thorndike (1924, p. 228) indicated this many years ago: 'All scientific measures of intelligence are measures of some product produced by the person or animal in question, or the way in which some product is produced.' In this sense function fluctuation is a performance fluctuation since we cannot measure mental functions directly. However, the use of a term like product or performance fluctuation encourages its possible confusion with crude variability, a confusion which Thouless very properly avoided by his use of the convenient 'fiction' or 'function' fluctuation.

It may be that Finney wishes to claim the concreteness of pseudo-behaviourism for his phrase, but this will only be at the expense of meaning. The justification for such 'mentalistic' concepts as 'function' is that they account most reasonably for observed facts (e.g. that the within-day reliability coefficients are markedly, or at least consistently, greater than the between-day coefficients), and they are hypotheses in Ayer's (1936) sense, and 'testable explanatory-cum-predictive assertions' in Ryle's phrase (1949, p. 25). Ryle defends mental concepts as explanations of behaviour characteristics. He suggests that 'when we describe people as exercising qualities of mind, we are not referring to occult episodes of which their overt acts and utterances are effects; we are referring to those overt acts and utterances themselves' (*ibid.* p. 25). Again: 'When a person is described by one or other of the intelligence-epithets such as "shrewd" or "silly", "prudent" or "imprudent", the description imputes to him not the knowledge, or ignorance, of this or that truth, but the ability, or inability, to do certain sorts of things' (*ibid.* p. 27). And finally: 'It is being maintained throughout this book that when we characterize people by mental predicates, we are not making untestable inferences to any ghostly processes occurring in streams of consciousness which we are debarred from visiting; we are describing the ways in which those people conduct parts of their predominantly public behaviour' (*ibid.* p. 51).

Burt (1950, p. 164) defends his phrase 'the inheritance of intelligence' in much the same way:

> But most psychologists...would nowadays be quite happy to accept the view of the logical positivists, and say that, in psychology as in other sciences, induction does not seek to prove general laws of causation, but merely to provide 'propositional functions'—logical models, in fact—from which verifiable singular propositions could be readily derived. Thus in claiming that intelligence is inherited, we are not to be understood as putting forward a new generalization about the 'concrete constitution' of the mind... but merely as formulating a compact and valuable scheme for particular predictions about particular individuals or particular groups of individuals.

And finally Tolman[1] (1932, p. 144) finds it necessary to introduce mental concepts into his system. He

> conceives mental processes as functional variables intervening between stimuli, initiating physiological states, and the general heredity and past training of the organism, on the one hand, and final resulting

[1] The author is indebted to Dr Thouless for this reference.

responses on the other. These intervening variables...are to be discovered, in the last analysis, by behaviour experiments. They have to be inferred 'back' from behaviour. They are precipitated out from the empirical correlations which can be observed between specific stimuli and initiating physiological states, on the one hand, and specific resultant acts, on the other. They are to behaviour as electrons, waves, or whatever it may be, are to the happenings in inorganic matter. There is nothing private or 'mentalistic' about them. They are pragmatically conceived, objective variables, the concepts of which can be altered and changed as proves most useful.

Finney also wishes to distinguish 'function fluctuation' from 'incentive fluctuation', and this yields some clue to his quarrel with Thouless's terminology. For Thouless the concept 'function fluctuation' is a description of a kind of behaviour and does not imply any psychological theory of the nature of mental functions or of their neurological correlates—it is what Cattell (1946) might term a descriptive 'surface' trait. The weather, state of health, strength of moods, incentives, boredom, each might be a source of fluctuation in different individuals on the same testing or the same individual on different testing occasions, either by serving to concentrate or deflect attention, or by surcharging or reducing the amount of mental energy available, or by any other means which need not necessarily be amenable to such metaphorical description. The term 'function fluctuation' implies neither its own determinants nor the mechanism by which these causal factors work. Finney, however, appears to imagine that the term implies that the cause of fluctuation lies within the function itself, which naturally fluctuates and would do so without, and over and above, the impact of any external or subjective stimulus. If function fluctuation were used in Finney's restricted sense, then it would be possible to distinguish between function fluctuation and incentive fluctuation, and indeed between function fluctuation and many other kinds, such as temperature fluctuations, mood fluctuation, health fluctuation, and so on. However, the term, as used more reasonably by Thouless, stands for a fact which could have any of the above causes.

II. EARLY METHODS OF MEASURING FUNCTION FLUCTUATION

The early methods of measuring function fluctuation stem clearly from a line of research which was concerned, quite mistakenly, with 'validating empirically' the Spearman-Brown 'prophecy' formula for predicting the increase in reliability to be expected from a given increase in the length of a test. It was rapidly discovered that the split-half coefficient, corrected for double length by the Spearman-Brown formula, yielded coefficients markedly, or at least consistently, higher than reliability estimates derived by correlating alternate forms of a test given on different testing occasions, and it was widely believed—Holzinger (1923*b*), Lanier (1927), Jordan (1935) and many studies occurring *passim* in the *Journal of Educational Psychology*—that this discrepancy was caused by the effect of function fluctuation, active between testing occasions, in reducing the alternate forms coefficients, while leaving unscathed the corrected split-half. While this may not be a completely valid explanation—there are spurious factors, notably that of the timing of the test (Thouless, 1939), which increase the size of the split-half and would do so even if functions did not fluctuate between tests—it did encourage psychometrists to believe that the discrepancy between perfect correlation and the alternate forms correlation could be considered as a measure of the activity of function fluctuation and test error, while the difference between unity and the split-half or corrected split-half might be

taken as a measure of the strength of test error alone. Accordingly, a comparison of the two coefficients might yield some measure of the influence of function fluctuation.

This is the position taken up by Paulsen (1931, p. 218):

> Present statistical theory does not anticipate any difference in the reliability coefficients calculated from measurements made all at one time or distributed over a period of time. However empirical results have shown that test-retest and the odd-even coefficients do differ markedly in size. The discrepancy indicates that the trait or function being measured is relatively 'variable'... changes in the performance of the individuals tested are responsible for the discrepancy between reliability coefficients based on adjacent trials (odd-even) and those separated by a time-interval of a week or a month (test-retest).

Paulsen formed the idea of correcting the test-retest coefficient for the attenuation due to test error, using the split-half reliabilities in the denominator of the attenuation formula, so that any discrepancy between the revised estimate and unity would measure the extent of the influence of function fluctuation. The revised estimate itself was to be termed a coefficient of trait variability (C.T.V.).

Quite apart from the criticism offered by Thouless that it would be more logical to use the difference between unity and the revised estimate as a C.T.V., it is doubtful whether the test-retest coefficient is to be preferred to the alternate forms estimate as a measure of between-occasions reliability, and the use of the split-half estimates as measures of test error appears inadvisable.

The idea of measuring the strength of function fluctuation by comparing between-day and within-day estimates of test reliability was carried a stage further by Dunlap (1933), who proposed to use Spearman's tetrad type of factorial analysis to determine whether 'quotidian variability', a term borrowed from Woodrow (1932) and identical with Paulsen's use of trait variability and Thouless's of function fluctuation, is active in the function being measured. 'When the tetrads from the four halves of two forms, each form given at a different time, equal zero, the evidence is that individual variability is negligible and that the Spearman-Brown formula will give exactly the same value as the correlation between two forms' (Dunlap, 1933, p. 448). Thouless has noted that the tetrad method is formally equivalent to his own, but he would prefer the safer course of using the within-day alternate forms correlation instead of Dunlap's split-halves as an estimate of test reliability affected by test error only.

Pursuing Dunlap's ideas, Ferguson (1941*a*) used Holzinger's bi-factor method of analysis to factorize the reliability coefficients obtained from M.H.T. 21, 23 and 26, which were administered to a sample of 212 children of 11+ in West Yorkshire schools at intervals of 1 week. The three tests were similar in content and regarded as parallel. Each was split into two (odd and even halves) and each half was correlated with every other. The matrix of intercorrelations then consisted of three correlations between halves of tests given on the same day, and twelve correlations between halves of tests given on different days. From the analysis of the matrix one general and three group factors emerged, each group factor clustering round the two halves of a test taken on the same day. From this Dunlap would have concluded that there was evidence for the activity of function fluctuation, but Ferguson notes that there is another possible explanation of the discrepancies between the within-day and between-day correlations—spurious factors such as the presumed correlation of test error and the speeding of the test. His conclusion —that either explanation is feasible and both together probable—demonstrates that his experiment has little psychological merit in that the experimental design, with its use of

split-halves, completely precludes the possibility of a decision between the two explanations which were appreciated by the author before the analysis had begun, and which had been evidently established and discussed in the research literature. Had Ferguson used parallel tests within any particular day as an estimate of test error instead of the split-half, and had the same results been obtained, then it might have been safely concluded that function fluctuation was active in the function under measurement.

The factorial method, originally suggested by Dunlap, seems not to have been used extensively in testing for function fluctuation. Yet its value to this end, especially when a large number of repetitions are available, is obvious. Consider the well-known decrease in the reliability coefficient with increase in time-interval between testing. R. L. Thorndike (1933), in his summary of the effect of the interval between test and retest on the constancy of the Stanford-Binet I.Q., demonstrated, in Ferguson's (1941 *b*, p. 125) phrase, 'that the correlation between Binet test and retest varies as some inverse function of the time-interval separating the successive testings'. Heim & Wallace (1949) administered Heim's A.H. 4 eight times to nine subjects and intercorrelated each day's scores with every other. The resulting matrix showed that the correlations calculated between scores on adjacent days are consistently higher than those calculated between scores on days farther apart. Thus the columns of correlations representing the middle testing occasions will show a smaller variance than those representing the earlier and later occasions, and the highest correlations always lie close to the diagonal cells of the matrix. Babington Smith (1950, p. 28) factorized this matrix, obtaining a large general factor and, inevitably, a small bipolar contrasting the first with the last four occasions. His conclusion runs: 'The first factor is heaviest on the middle tests, not, I suggest, because there is a factor of intelligence which is most strongly represented there, but because the situation of taking a test has undergone progressive change, and the middle tests are not so different from any other test as a test at one end is from a test at the other end. The loadings on the second, bipolar, factor are associated with the order in which the tests are given and the progressive changes.' However, if we see the matrix as contrasting within- and between-adjacent days' correlations, then Babington Smith's 'progressive change in the test situation' becomes the progressively active influence of function fluctuation or perhaps a time-trend with fluctuations, and the explanation of the term 'different' in the phrase, 'the middle tests are not so different from any other tests', is that functional variability has not had time to act so extensively in the middle tests. He suggests (*ibid.* p. 79) that 'the illustration [Heim's data] suits me because it is one of the few cases I know where we have material which will bring out the effects of serial links on the factor-structure'. Rather the data appears to be not a criticism of factor-analysis but an illustration of one of its possible uses. The fact that the matrix contains evidence of function fluctuation which factor-analysis records is a ground for faith in the value of the factorial method.

A caveat should be entered here against the validity of any interpretation of the second and third factors, which are probably insignificant since so few people took the tests. Babington Smith does not appear to have tested the significance of each residual matrix, and does not explain why he stopped factorizing when he did. Should the last two factors prove to be insignificant, much of the point of each interpretation is lost. However, this would not affect the present writer's case since the decrease in the reliability coefficient with increased time-interval between testing is a well-established psychometric fact, and

tests with larger numbers of subjects would doubtless show the same pattern of intercorrelations as that presented by Heim.

Of course it may be agreed that the test-retest correlation is not a reliability coefficient in that there is the possibility of direct memory transference which encourages a correlation of test error. Kelley (1924, p. 202) says of the test-retest: 'We may call such a situation one in which there is a correlation between errors, meaning that whatever elements of uncertainty or chance operated in the solution of the first question, this would tend to operate in the same manner in the solution of the second.' However, against this can be placed the fact that any correlation of test error affects all the coefficients and may not, therefore, be used as an explanation of differences between them. Functional variability or rather functional trend is the most reasonable explanation of the pattern of Heim and Wallace's correlations, although the design of their experiment is such that no definite proof is afforded. In any case it seems clear that factorial analysis could be a useful method of testing for evidence of function fluctuation if parallel forms of a test were administered on each testing occasion so that a reasonable estimate of the extent of test error could be made.

Woodrow (1932), in trying to test for the existence of what he called 'quotidian variability', substituted for the correlation, with its logic of similarity, the standard deviation with its logic of differences, and suggested, without apparently being aware of the then-novel method of analysis of variance, that a comparison of the extent of variability between testing occasions to the average variability within testing occasions might yield a good index of 'quotidian variability' for a subject repeating the test. Thouless has demonstrated how Woodrow's clumsier statistic can be replaced by the more elegant analysis of variance.

III. Thouless's correlation of differences

Thouless derived his correlational method of testing for evidence of function fluctuation and measuring its amount from early British psychometric research, specifically Spearman's (1904*a*, *b*, 1910) attempt to deal with the problem of errors of measurement in psychological data, errors which attenuate the hypothetical 'true' correlation between true error-free scores on the correlated variables. Spearman (1904*a*, p. 88) divides the errors into two groups: 'accidental' or 'compensating' and 'systematic', 'constant', or 'non-compensating'; and proposes to correct for the attenuation, induced by the former, by a formula which necessitates the administration of two independent tests of the same mental function together on at least two different testing occasions, 'on the idea that the size of these accidental errors can be measured by the size of the discrepancies between different measurements of the same things' (Spearman, 1910, p. 274).

Starting from this background, Thouless applied this double-test-retest method to a more useful purpose. He assumes that two alternate and parallel forms of a test, A and B, are administered on two different testing occasions. Each individual has, therefore, four scores, A_1 and B_1 on the first testing occasion, and A_2 and B_2 on the next. Thus six correlations may be calculated: the within-day $r_{A_1B_1}$ and $r_{A_2B_2}$, between-day test-retest $r_{A_1A_2}$ and $r_{B_1B_2}$, and between-day alternate forms $r_{A_1B_2}$ and $r_{B_1A_2}$. Thouless states that, if test error is the only source of variation, and if errors of measurement are not correlated, then there will be no significant correlation of the difference between the test and retest scores for each of the two tests, i.e. $r_{(A_1-A_2)(B_1-B_2)}$ will not differ significantly from zero.

If, however, function fluctuation is active, then the scores of the subjects in the retesting will deviate positively or negatively, the extent and the direction of the deviations being common to A_2 and B_2. This implies that the within-day reliabilities, $r_{A_1B_2}$ and $r_{A_1B_2}$, have a common factor of 'temporal contiguity' (Thouless's term) which tends to increase their separate correlations, while relatively the alternate forms $r_{A_1B_2}$ and $r_{B_1A_2}$ show a decreased relationship. The amount $r_{(A_1-A_2)(B_1-B_2)}$, which may be termed the 'criterion' of whether function fluctuation is present, suffers from the fact that it is not independent of the size of r_{AB}, and so is not suitable for indicating the amount of fluctuation present. Accordingly, Thouless derives an index of that amount by dividing the criterion by the average of the within-day reliabilities

$$\frac{r_{(A_1-A_2)(B_1-B_2)}}{\frac{1}{2}(r_{A_1B_1}+r_{A_2B_2})}.$$

This is termed the 'double test-retest index of function fluctuation' (Thouless, 1936, p. 333), and will be designated I_1.

The method was earlier used by Brown (1910) in his attempt to disprove the assumption of no error-correlation which Yule had demonstrated to be the basis of Spearman's original and most famous correction for attenuation. Brown & Thomson (1940, p. 158) argue, however, that any significant correlation of differences is best taken as evidence not of error correlation but of 'functional variability', and Holzinger (1923*a*), using the same method, concurred with this view.

The present writer has been able to trace few researches making use of Thouless's proposals. Vernon (1937) stressed their importance. Ferguson (1941*b*) gives an account of the method but is mistaken in his assertion that a test of significance, presumably the normal standard error to the correlation, is applicable to the index. Jackson & Ferguson (1941) suggest that there is no evidence for function fluctuation with cognitive abilities without specifying any research. Connor (1952) uses the Spearman tetrad type of analysis. Remmers & Whisler (1938) note Thouless's formulation, and attempt to improve on it. They take tests U and V as measures of the same function, each being further subdivided into items, U_1, ..., U_n and V_1, ..., V_n, and these items so arranged that there is no constant time-difference beteeen the items of the different series—$U_1V_1U_2V_2$ or $U_1U_2V_1V_2$. They assert that the '"true" relation between U and V is given by the correction for attenuation' (Remmers & Whisler, 1938, p. 90). If all the V items are now formed into one test and all the U items into another, then r_{UV} corrected for attenuation may be computed, and the discrepancy between the two corrected correlations or between their squares may be taken as a criterion of whether function fluctuation is present. Quite apart from the fact that this method offers no advantages over Thouless's, the validity of the formula for correction for attenuation in this context is doubtful (Thouless 1939).

One point about Thouless's account appears to have gone without discussion. Thouless suggests that the index will have unity as its maximum value. Speaking of the index, he writes (Thouless, 1939, p. 333): 'Its value is zero if there is no day-to-day variation of the mental function measured by the two tests. Function fluctuation produces a positive value of the index which would be unity if function fluctuation were the sole cause of absence of complete correlation between repetition of the tests (i.e. if there were no test unreliability).' However, there is no apparent mathematical reason why the maximal

value of the index should be unity, although in practice it will probably be less than unity with tests of high reliability, and the present writer has found that tests showing $r_{AB} < 0{\cdot}3$ sometimes yield indices greater than unity.

IV. Finney's analysis of variance model

The use of analysis of variance in testing for function fluctuation in the repeated performances of an individual was first suggested by Thouless. Finney (1953) has recently suggested a model for use with groups of subjects. Making the usual assumptions, and provided each test on each occasion is independently standardized to the same mean and standard deviation, with q subjects being given the same set of p tests on each of r occasions, for all functions the analysis will be as shown in Table 1. The test for evidence

Table 1. *Analysis of variance model*

Variations	D.F.	Observed mean square
Subjects (S)	$q-1$	S
Tests (T)	0	—
Occasions (O)	0	—
ST	$(p-1)(q-1)$	A
SO	$(q-1)(r-1)$	B
TO	0	—
STO	$(p-1)(q-1)(r-1)$	C
Total	$pr(q-1)$	

of function fluctuation is the variance ratio $F = B/C$, but, like Thouless's criterion with which it is directly correlated, it varies with the size of r_{AB}. Finney cites three indices, two of which have value for the present research. His second index (I_2) is the ratio of the fluctuation variance to the error variance, estimated by $(B/C-1)/p$. This will have little relation to Thouless's (I_1) because it is perfectly correlated with the variance-ratio, but it has the merit of serving as a check on I_1. It may be shown[1] that Thouless's criterion is approximately equal to

$$\frac{\dfrac{B-C}{p}}{\dfrac{B-C}{p}+C}\,\frac{\dfrac{1}{p}\left(\dfrac{B}{C}-1\right)}{\dfrac{1}{p}\left(\dfrac{B}{C}-1\right)}+1=\frac{I_2}{I_2+1}.$$

These equivalences are designated approximate—Thouless's criterion is not exactly equal to $I_2/(I_2+1)$, mainly because of the heterogeneity of the standard deviations used in the calculation of the correlations.

Finney's third index (I_3) is the ratio of the fluctuation variance to the variability of subjects, estimated by

$$\frac{r\,(B-C)}{S-A-B+C}.$$

This appears to be psychologically akin to I_1. However, since it is a ratio with many terms, the size of its standard error will be considerable. Moreover, since it contains terms other than B and C, there is no statistical necessity that it should be significant even when its related criterion is significant, whereas it is psychologically fundamental that a signifi-

[1] These approximate equivalences were suggested to the present writer by Miss V. R. Cane of the Medical Research Council Applied Psychology Unit, Cambridge, and by Dr H. E. Daniels, of the Statistical Laboratory, the University of Cambridge.

cant criterion should give rise to a significant index. Both points may be illustrated from the approximate standard error to I_3 given to the present writer by Dr Daniels.

$$\text{S.E.}_{I_2}=\frac{y\,(y+r)}{r}\sqrt{\left[\frac{2}{k}\left\{\frac{S^2+A^2}{(S-A)^2}+\frac{(B^2+C^2)}{(B-C)^2}\right\}\right]},$$

where $y=I_3$, $k=$number of subjects. Table 2 shows relevant data for three selected functions concerning the size of B/C, the estimate of I_3 and the standard error of I_3. Thus each of the variance ratios B/C is significant at the 5% level (1·48) while not one of the

Table 2. *The significance of the criterion and related indices*

Functions	B/C	I_3	S.E.$_{I_3}$
26	2·52	1·916	1·5
25	1·594	1·087	1·67
21	2·445	11·79	43·57

comparable indices is significant. Accordingly, it seems wise to restrict the use of the index to its raw form and to compare indices by ranking only.

V. EXPERIMENTAL STUDY

Parallel tests of twenty-six functions were administered on the same day to a group of seventy-four children in the scholarship class in two schools in Cambridgeshire, and re-administered approximately 1 week later. Each of the four distributions of scores for any function was normalized to the same mean and standard deviation, and Table 3 gives the comparative statistics for Thouless's and Finney's models.

It is clear that the two models give much the same results. However, it is obvious that I_3 has the disadvantage of being occasionally negative. This may happen when the fluctuation variance is smaller than the error variance, especially in the case of tests showing high reliability and discriminating well between objects. Such a negative index would quite validly represent—at least from the psychological point of view—a low estimate of the amount of fluctuation present. However, a negative index may also derive from the fact that in the denominator $A+B>S+C$. This condition—responsible for the negative I_3 of functions 7 and 22—is a product of the joint effects of poor test discrimination between subjects (relatively low S variance), marked differential reaction by the subjects to the different tests (relatively large ST variance) and a greater fluctuation than error variance ($B>C$). Now in Table 3, the negative I_3 indices for functions 7 and 22 are negatively correlated with their corresponding I_1 indices, which are relatively large and positive. While there is no reason to expect the rank-order coefficient between I_1 and I_3 to be unity always (since the one contains terms that the other does not), yet it should be significantly positive and high.

The solution to the problem may be that I_3 is only a good index under certain conditions. Consider it in the form

$$I_3=\frac{\frac{B}{C}-1}{\frac{S}{C}-\frac{A}{C}-\frac{B}{C}+1}.$$

If sets of random numbers rather than test scores had been chosen, the mean squares would arise simply by chance, and the ratios S/C, A/C and B/C would approximate to

unity. In this case I_3 tends to approach the indeterminate form O/O which is highly unstable, and the smallest change in many of the mean squares could swing the index from minus infinity to plus infinity. The conclusion is, therefore, that I_3 is only of service if B/C is significant and A/C insignificant, assuming that S/C is also significant and that the tests do actually discriminate between pupils. The index I_3 may be a good measure

Table 3. *Comparative statistics from Thouless's and Finney's models*

Function	Thouless's criterion	I_1	$F=B/C$	I_2	$\frac{I_2}{I_2+1}$	I_3
1	0·471[1]	0·512	2·752	0·876	0·462	0·0475
2	0·442[1]	0·526	2·617	0·8085	0·442	0·089
3	0·400[1]	0·435	2·384	0·692	0·408	0·024
4	0·143	0·159	1·291	0·1455	0·127	0·011
5	0·410[1]	0·500	2·374	0·687	0·407	0·078
6	0·526[1]	0·715	3·252	1·126	0·530	0·314
7	0·154	1·247	1·337	1·685	0·144	−5·488
8	0·275[1]	0·658	1·737	0·3685	0·269	0·492
9	0·498[1]	0·868	2·98	0·99	0·497	1·331
10	0·206	0·379	1·549	0·2745	0·215	0·192
11	0·549[1]	0·692	3·571	1·2855	0·562	0·366
12	0·178	0·335	1·47	0·235	0·190	0·224
13	0·283[1]	0·590	1·862	0·431	0·301	0·486
14	0·374[1]	0·900	2·247	0·6235	0·384	0·937
15	0·371[1]	0·672	2·144	0·572	0·364	0·662
16	0·195	0·424	1·454	0·227	0·185	0·454
17	0·392[1]	0·806	2·271	0·6355	0·388	1·426
18	0·246[1]	2·149	1·667	0·3335	0·250	1·362
19	0·347[1]	1·209	1·974	0·487	0·328	2·528
20	0·090	0·548	1·227	0·1135	0·102	1·394
21	0·411[1]	1·137	2·445	0·7225	0·419	11·79
22	0·162	1·296	1·364	0·182	0·154	−1·775
23	0·210	0·520	1·503	0·2515	0·201	0·532
24	0·018	0·120	1·068	0·034	0·033	0·193
25	0·242[1]	0·785	1·594	0·297	0·229	1·087
26	0·409[1]	0·863	2·52	0·760	0·432	1·916

[1] Significance at the 5% level.

Table 4. *Correlations for functions* 7 *and* 22

Function	A_1B_1	A_2B_2	A_1A_2	B_1B_2	A_1B_2	B_1A_2
7	0·208	0·039	0·291	0·205	0·014	0·002
22	0·152	0·098	0·437	0·202	0·116	−0·083

of the extent of function fluctuation when it is certain that only one function is being measured, and the function fluctuation is significant. This hypothesis is borne out by a glance at the functions in Table 3, where B/C is not significant or of borderline significance (7, 10, 12, 16, 20, 22, and 23), and by considering the correlations between the test administrations, in which the tests (non-cognitive) are obviously not parallel.

The instability of I_3 can also be illustrated from the S.E. provided by Dr Daniels, in which, whenever $S=A$ or $B=C$, the S.E. moves quickly to infinity. It has been suggested to the present writer[1] that the form of the S.E. condemns I_3 as a suitable index.

In general, if the tests are not really parallel and if the activity of function fluctuation is not significant by the B/C test, I_3 loses precision, becomes unstable and can acquire any value, positive or negative, without meaning anything.

[1] By J. S. Macpherson, who gave valuable advice about this passage.

VI. Conclusion

The best simple methods of testing for function fluctuation, those of Thouless and Finney, give similar results, and it is suggested that they should be used in future research into the amount of fluctuation present in cognitive and non-cognitive functions. The present writer is carrying out such a programme with scholarship and pre-scholarship children, and research with different tests of many functions should be extended to other populations.

I am greatly indebted to Dr R. H. Thouless for his critical supervision of this paper, especially his pruning of the vaguer expressions. Ben S. Morris, Director of the National Foundation for Educational Research in England and Wales, suggested the topic of research, gifted some of the testing material, allowed me to consult Finney's memorandum, and has shown great interest in the progress of the work. John S. Macpherson, Research Officer of the Scottish Council for Research in Education, gave valuable linguistic and statistical advice. Finally, I must thank the Carnegie Trust for a grant in aid of the research.

References

Ayer, A. J. (1936). *Language, Truth and Logic.* London: Gollancz.

Brown, W. (1910). Some experimental results in the correlation of mental abilities. *Brit. J. Psychol.* III, 296–322.

Brown, W. & Thomson, G. H. (1940). *The Essentials of Mental Measurement.* Cambridge University Press.

Burt, C. (1950). The trend of national intelligence. *Brit. J. Sociol.* I, 154–68.

Cattell, R. B. (1946). *The Description and Measurement of Personality.* London: Harrap.

Connor, D. V. (1952). The effect of temperamental traits upon the group intelligence test performance of children. Unpublished Ph.D. thesis, University of London.

Dunlap, J. W. (1933). Comparable tests and reliability. *J. Educ. Psychol.* XXIV, 442–53.

Ferguson, G. A. (1941*a*). A bi-factor analysis of reliability coefficients. *Brit. J. Psychol.* XXXI, 172–82.

Ferguson, G. A. (1941*b*). *The Reliability of Mental Tests.* London: U.L.P.

Finney, D. J. (1953). *A Note on the Measurement of Performance Fluctuation.* Memorandum to the National Foundation for Educational Research in England and Wales.

Guilford, J. P. (1950). *Fundamental Statistics in Psychology and Education.* New York: McGraw Hill.

Gulliksen, H. (1950). *The Theory of Mental Tests.* New York: Wiley.

Heim, A. W. & Wallace, J. G. (1949). The effects of repeatedly testing the same group on the same intelligence test. Part I. Normal subjects. *Quart. J. Exp. Psychol.* I, 151–9.

Hertzman, M. (1938). Two equations for the study of variability. *Amer. J. Psychol.* LI, 571–4.

Holzinger, K. J. (1923*a*). An analysis of errors in mental measurement. *J. Educ. Psychol.* XIV, 278–88.

Holzinger, K. J. (1923*b*). Note on the use of Spearman's prophecy formula for reliability. *J. Educ. Psychol.* XIV, 302–5.

Hull, C. L. (1927). Variability in amount of different traits possessed by the individual. *J. Educ. Psychol.* XVIII, 97–106.

Jackson, R. W. B. & Ferguson, G. A. (1941). *Studies on the Reliability of Tests.* Toronto: Bulletin 12. Department of Educational Research, the University of Toronto.

Jordan, R. C. (1935). An empirical study of the reliability coefficient. *J. Educ. Psychol.* XXVI, 416–26.

Kelley, T. (1924). *Statistical Method.* New York: Macmillan.

Lanier, L. H. (1927). Prediction of the reliability of mental tests and tests of special abilities. *J. Exp. Psychol.* X, 69–113.

Paulsen, G. B. (1931). A coefficient of trait variability. *Psychol. Bull.* XXVIII, 218–19.

Preston, M. G. (1940). Concerning the determination of trait variability. *Psychometrika,* V, 275–81.

Remmers, H. H. & Whisler, L. (1938). Test reliability as a function of method of computation. *J. Educ. Psychol.* XXIX, 81–92.

Ryle, G. (1949). *The Concept of Mind.* London: Hutchinson.

Smith, Babington (1950). Factor analysis. Part II. An evaluation of factor analysis from the point of view of a psychologist. *J. R. Statist. Soc.* B, XII, 73–85.

Spearman, C. S. (1904*a*). The proof and measurement of association between two things. *Amer. J. Psychol.* xv, 72–101.

Spearman, C. S. (1904*b*). 'General Intelligence' objectively determined and measured. *Amer. J. Psychol.* xv, 201–92.

Spearman, C. S. (1910). Correlation calculated with faulty data. *Brit. J. Psychol.* iii, 271–95.

Thorndike, E. L. (1924). The measurement of intelligence. *Psychol. Bull.* xxxi, 219–52.

Thorndike, R. L. (1933). The effect of the interval between test and retest on the constancy of I.Q. *J. Educ. Psychol.* xxiv, 543–9.

Thouless, R. H. (1936). Test unreliability and function fluctuation. *Brit. J. Psychol.* xxvi, 325–43.

Thouless, R. H. (1939). The effect of errors of measurement on correlation coefficients. *Brit. J. Psychol.* xxix, 383–403.

Tolman, E. C. (1932). *Purposive Behaviour in Animals and Men.* New York: The Century Co.

Vernon, P. E. (1937). The Stanford-Binet Test as a psychometric instrument. *Character & Pers.* vi, 99–113.

Woodrow, H. (1932). Quotidian variability. *Psychol. Rev.* xxxix, 245–56.

(*Manuscript received* 12 *April* 1954)

BREADTH OF EXPERIENCE, EASE OF DISCRIMINATION, AND EFFICIENCY OF GENERALIZATION

By WILLIAM BEVAN and PER SAUGSTAD

Emory University, Georgia, and the University of Oslo[1]

I. *Introduction* (p. 13). II. *Experimental design and specification of hypotheses* (pp. 13–15). III. *Experimental situation* (pp. 15–16): (i) *Materials* (p. 15); (ii) *Experimental groups* (p. 15); (iii) *Experimental sessions* (pp. 15–16); (iv) *Measures of performance efficiency* (p. 16). IV. *Results* (pp. 16–17). V. *Discussion* (pp. 17–19): (i) *Evaluation of Hypothesis* (pp. 17–18); (ii) *Subsidiary comment* (pp. 18–19). VI. *Summary* (p. 19). *References* (p. 19).

I. Introduction

Most psychological theories to the present appear to possess in common at least one important untested assumption: An organism's discriminatory mechanisms are regarded to exist as given, essentially functionally intact and ready for use, as part of its biological equipment. This has had obvious consequences. The question of representativeness of design does not present itself; the significance, for general theory, of truly developmental studies and studies of individual differences, is usually underestimated; and the adequacy of control of the subject's pre-experimental past experience—specification as a rule is limited to statements about age and *naïveté* with regard to the experiment's purpose—is generally not seriously considered.

Recently reported data (e.g. Brunswik's (1939) on the effect of varying the relative frequency of required responses; Helson's (1948) and Johnson's (1944) on the effect of varying the relative frequency of various magnitudes of stimulation; Riesen's (1950) on arrested vision in chimpanzees, Ames' (1946) on visual illusions; and Harlow's (1949) on interproblem transfer) meanwhile support the view that mature discriminative behaviour requires not only establishing some sort of functional juncture between the action of the discriminatory mechanism and that involved in the overt expression of behaviour but also the sufficient development of the discriminatory mechanism itself (cf. e.g. Hebb's (1949) developmental theory of perception, Cantril's (1950) concept of assumptive world and Harlow's (1949) of learning set). Yet little has been done, with the exception of Helson's (1948) studies of intra-organismic reference standards, to specify empirically the parameters of such a developing mechanism. Work such as Harlow's, however, suggests that there are dimensions other than those traditionally recognized that may be fruitfully explored. In a discussion of problem-solving Harlow (1951, p. 469) has commented: 'Breadth, rather than intensity of training, appears to be the key...whether we deal with subhuman or with human subjects.'

Accordingly, the aim of the present investigation has been to attempt to create a situation in which breadth of experience (training) may be conveniently given objective definition and its relation to performance efficiency in a 'new' situation evaluated.

II. Experimental design and specification of hypotheses

The plan of the experiment consists of a combination of two traditional designs: that of the method of constant stimuli for measuring judgemental acuity and that of the transfer

[1] The present work, done at the University of Oslo during the 1952–3 academic year, represents the joint effort of both writers while the former was in residence as a Fulbright research scholar.

of training experiment. The subjects' task involved making differential judgements with one of several ranges of stimuli about a common standard on four consecutive days and then on the fifth with a non-overlapping test range. Breadth of experience is considered defined by the magnitude of the training range, novelty of situation by the difference in position of training and test ranges on the stimulus continuum. Performance efficiency is expressed in terms of the accuracy and precision of the mean judgements made with the test range.

Since no concern was directed toward the criterion of representativeness of design, the experimenters have no wish to imply that the unidimensional definition they have applied to breadth of experience is adequate for the 'normal' condition for which one may ultimately wish to predict. Nor do they claim that the relatively invariate situation they arranged for its operation have yielded data which in any crucial respect are typical of 'normal conditions'. Representative complexity was, possibly, sacrificed to gain ease of objective specification for the key concepts.

Thus the present experimental task may lack the novelty that the laboratory discrimination problem is presumed to have for the traditionally assumed-to-be naïve white rat, dog, or rhesus monkey, for all of our subjects had already developed, probably to a highly refined degree, a concept of stimulation differential. We have rather *arbitrarily* selected the lifted-weight task as our discrimination problem, mainly because we suspect that it provides a situation requiring formal judgement met less frequently by our subjects in their pre-experiment past. We assumed, furthermore, that our subjects' skill at judging had not been developed to a degree such that it could not be further improved by the training procedures of our experiment. We also assumed, of course, that prior to this training subjects randomly grouped were on the average equally sophisticated. Harlow's conclusion quoted above was consequently transformed as follows: If subjects be allowed experience (training trials) in judging the relative heaviness of lifted weights, equal in frequency but varying in range, those with wider practice will, when confronted by weights outside the training range, be more accurate and/or precise in their discriminations. One or both types of subject should be superior in performance to those receiving no experimental training prior to presentation of the test range.

If this first hypothesis were confirmed, a second was generated from the Neo-Pavlovian theory of generalization gradients (Hull, 1943) to account for the superiority of the broader training. If subjects be allowed training experience, equal in amount and range, but differing in its similarity to the test experience (adjacency of training and test ranges on the stimulus continuum), those receiving the more similar experience will display greater accuracy and/or precision in their performance of the test task.

Should the second hypothesis lack confirmation, another means of explaining the superiority of the broader training would have to be sought. The stimulus series used for narrow and broad training differ not only in range about a common central magnitude but in the size of the inter-weight interval. This suggests that degree of distinctness of the individual training experiences is important for generalization and is reminiscent of the Köhler & von Restorff report (1933) on the relationship of 'crowding' to ease of recall, except that, of course, the Köhler & von Restorff 'crowded' and 'isolated' materials were qualitatively different. That inter-weight interval may be important is also suggested by the human-engineering studies (Chapanis, Garner & Morgan, 1949) on interval size and efficiency of dial reading. A final hypothesis was therefore formulated: If subjects be allowed training experience equal in total amount and derived from the same range

within the stimulus continuum, the magnitude of the standard-variable interval in the training series will prove to be a significant factor in the accuracy and/or preciseness of their test performance. In the present instance the superiority of the wider interval training group is predicted.

III. EXPERIMENTAL SITUATION

(i) *Materials*

The weights employed as stimulus material were opaque plastic cylinders 5 cm. in diameter and 8 cm. high filled with a combination of lead shot, saw-dust and cotton. They were made up in four training series and a test series. Training series A, the long-range, wide-interval series, consisted of five weights, each weight differing from the next heaviest in the series by 40 g. and with the middlemost serving as the standard: 220 g., 260 g., 300 g. (S), 340 g., 380 g. Training series B, the short-range, narrow-interval series, like series A consisted of five weights with the standard at 300 g. The interval between adjacent weights of 10 g., however, effected a range and interval size one-fourth that of series A: 280 g., 290 g., 300 g. (S), 310 g., 320 g. Training series C, also a short-range, narrow-interval series, is the adjacent series. Both its lower end and its midpoint are closer to those of the test series than are the corresponding magnitudes in series A and B. Its interval size is proportionately equal to that of series B: 205 g., 212·5 g., 220 g. (S), 227·5 g., 235 g. Training series D is the long-range, narrow-interval series. Constructed about a midpoint of 300 g., its range is equal to that of series A, its standard-variable differential never exceeding the maximum differential of series B. This was accomplished by arranging sixteen weights into five overlapping short-range, narrow-interval sub-series: 220 g., 230 g., 240 g. (S), 250 g., 260 g.; 250 g., 260 g., 270 g. (S), 280 g., 290 g.; 280 g., 290 g., 300 g. (S), 310 g., 320 g.; 310 g., 320 g., 330 g. (S), 340 g., 350 g.; 340 g., 350 g., 360 g. (S), 370 g., 380 g. Finally the test series consisted of nine weights 4 g. apart about a standard of 100 g.: 84 g., 88 g., 92 g., 96 g., 100 g. (S), 104 g., 108 g., 112 g., 116 g.

A three-sided, roughly 'Z' shaped wooden screen, with side panels 40 cm. by 50 cm. and middle panel 50 cm. square, and with a semicircular aperture through which the subject's lifting hand could be inserted at the bottom centre of the middle panel, was devised in order to eliminate visual cues during testing. The weights were presented on a rubber-backed felt pad to minimize the possibility of auditory cues.

(ii) *Experimental groups*

Ninety-six persons, seventy-nine medical and seventeen natural science students at the University of Oslo, all in their early twenties, served as subjects. None had had experience in or knowledge of weight-lifting experiments prior to his participation. Each subject was randomly assigned to one of five groups. Those assigned to Group 1 attended only one experimental session during which they were presented with the test series. Members of the other four groups received training in weight discrimination with one of the four training series prior to test. The unequal numbers in the groups resulted from a few subjects withdrawing their participation before completion of the training sequence.

(iii) *Experimental sessions*

The training sessions, which required about 15 min. each, were held on four consecutive days. They were followed on the fifth day by the test session, also about 15 min. in duration. During each session each subject made eighty comparative judgements (Group 5,

which received training series D made eighty-two on each training day), half with the standard presented first, half with the order reversed, in random sequence. The random alternation of the temporal order in which the standard was presented on successive trials was, of course, executed with the hope of distributing the bias contributed by the psychophysical time error. During the training sessions the standard (standards in the case of series D) was presented with itself as frequently as with each other weight. In the test series it was never presented with itself. Subjects, all trained and tested individually, were requested to lift with the non-dominant hand in an effort to reduce to some extent the effect of weight-experience gained before the experiment. The testing procedure, that followed in the method of constant stimuli, is described in the following instructions to the subject:

This is an experiment in perception which aims at demonstrating how the ability to form judgements of a certain type develops. For this purpose we shall ask you to judge the weight of various objects.

Which is your dominant hand? Put the *non-dominant* hand through the aperture of the screen before you. The wrist ought to be well inside the aperture. Now I will present to you a number of objects which are alike in every respect except their weight. You will be given one object at a time. When you feel the object against the tip of your fingers grasp it as you would a glass.

With regard to the manner in which you are to lift the object, there are certain matters to which you must pay careful attention. Lift every object only once, raising it quickly and putting it down immediately. Then give your judgement. When you lift, the whole arm from the wrist up should rest on the table; in other words, you are to lift only from the wrist. Be careful that you raise the whole hand and do not rest any of your fingers on the table while you are lifting.

You will be given pairs of objects one after the other. When you have lifted both, you are to tell if you think one is heavier or lighter than the first. If the last one is lighter, you say, 'light'; if it is heavier, you say, 'heavy'.

(iv) *Measures of performance efficiency*

The relative frequencies of 'heavier than' judgements for each standard-variable comparison of the test series were first combined by means of Spearman's summation method (Woodworth, 1938) to yield a mean heaviness score and standard deviation for each subject. The cut-tail correction was applied when appropriate. The mean heaviness score is comparable (assuming a normal distribution of responses) to the P.S.E. gained graphically and provides information on the degree of correspondence between the objective midpoint of the stimulus range and its subjective equivalent. It is taken as an *index of accuracy*. The standard deviation provides information about the consistency of a subject's performance about his own mean. It is taken as an *index of preciseness*. After these indices had been derived for each subject, they were used as raw scores in group comparisons. The indices entered in columns 3 and 4 of Table 1 are thus arithmetic averages of the individual accuracy and precision indices. A third proficiency score, the *index of uniformity*, was also used. This is the standard deviation of the *mean* accuracy index and indicates the extent to which subjects within each experimental group are alike in accuracy.

IV. Results

Table 1 presents the mean indices of accuracy and preciseness as well as the uniformity scores for the several experimental groups; Table 2 presents the extent to which the group means deviated from perfect accuracy; Table 3 presents the results of the various possible intergroup comparisons of preciseness and uniformity. Examination of Table 1, column 3, and Table 2 indicates almost perfect accuracy in all groups except Group 4, that trained on the adjacent series, which shows a small but significant tendency toward underestimation. The performance of Group 2, trained on the long-range, wide-interval

series is reliably more precise than that of any of the other training groups. These last-mentioned, meanwhile, do not differ reliably from those receiving no formal practice before the test. Finally, there are clear differences among the groups in uniformity. As in the case of precision, Group 2 is the superior group. Groups 1, 4 and 5 do not differ significantly among themselves. Group 3, the short-range, narrow-interval group, is inferior to all other groups.

Table 1. *Mean indices of accuracy and preciseness, index of uniformity expressed in grams for the several experimental groups*

Experimental group	No. subj.	Mean index of accuracy	Mean index of preciseness	Index of uniformity
(1) No training	21	100·15	11·66	2·51
(2) Training series A	19	99·94	9·18	0·97
(3) Training series B	19	100·31	11·96	4·64
(4) Training series C	20	101·14	10·92	2·29
(5) Training series D	17	100·60	12·17	2·45

Table 2. *Summary of 't' tests to determine if experimental groups differed reliably from perfect accuracy*

	Difference between mean accuracy index and perfect accuracy for each group	t	P
(1)	0·15	0·27	>0·05
(2)	0·06	0·26	>0·05
(3)	0·31	0·26	>0·05
(4)	1·14	2·19	<0·05
(5)	0·60	0·98	>0·05

Table 3. *Summary of intergroup comparison with regard to preciseness and uniformity*

Groups compared	D.F.	Diff.	t	P	Diff.	t	P
1–2	39	2·48	3·39	<0·01	1·54	3·67	<0·01
1–3	39	0·30	0·39	>0·05	2·13	2·45	<0·05
1–4	40	0·74	0·87	>0·05	0·22	0·41	>0·05
1–5	37	0·51	0·63	>0·05	0·06	0·11	>0·05
2–3	37	2·78	3·66	<0·01	3·67	4·58	<0·01
2–4	38	1·74	2·07	<0·05	1·32	3·22	<0·01
2–5	35	2·99	3·74	<0·01	1·48	3·22	<0·01
3–4	38	1·04	1·19	>0·05	2·35	2·76	<0·01
3–5	35	0·21	0·25	>0·05	2·19	2·49	<0·05
4–5	36	1·25	1·38	>0·05	0·16	0·28	>0·05

V. DISCUSSION

(i) *Evaluation of hypothesis*

With these data in mind, we may review and evaluate the status of the hypotheses specified above. The greater preciseness of Group 2 as compared to Group 3 and their lack of difference in accuracy add up to partial confirmation of the first hypothesis. This confirmation is strengthened by the greater intra-group uniformity of Group 2. Comparison of the performance of these groups with that of Group 1 indicates that the superiority of Group 2 represents a facilitating effect of this type of training. The failure of Group 3's performance to surpass that of the no-training group (Group 1) in precision, and the high-level accuracy of all groups suggests the extent to which *pre-experimental* heaviness experience benefits judging in the present experimental situation. If practice is required to match the performance of Group 3, it could have been provided for the no-training group only during the test session itself. The superiority of Group 2 over

Group 3, assuming it not to reflect sampling error, meanwhile indicates that even with relatively sophisticated subjects, types of training may be employed which will effect greater proficiency in certain aspects of performance. The relative lack of uniformity in Group 3 meanwhile suggests that some kinds of training may interfere with rather than facilitate the operation of the discriminatory mechanism.

Comparison of Group 3 with Group 4 reveals no significant difference in preciseness but the inferiority of Group 4 in accuracy. Meanwhile Group 2 retains its superiority in both preciseness and uniformity. This means a rejection of our second hypothesis and points to the inadequacy of the similarity theory as it is usually defined for discrimination experiments of this sort. The excellence of the long-range training group (Group 2) cannot be attributed to the fact that the training series contains weights which are closer in the stimulus continuum to the test series than are corresponding items of the short-range series.

Finally, comparison of Groups 2 and 5 provides a test of our third hypothesis. The significantly smaller average preciseness score and uniformity score mean confirmation with regard to these criteria. At the same time they mean a rejection of Harlow's 'breadth hypothesis' as we have specified it for this experiment. Group 2 is superior, not because its pre-test weight lifting experience covered a wider range of weights, but because it included paired items more widely separated on the weight continuum and thus more readily discriminable. Breadth of experience or breadth of training, should this experiment be representative of the requirements of at least the simplest discrimination problems, is most adequately defined, not in terms of total range of magnitudes presented to the subject, but in terms of the range of differences in magnitudes provided in the sequence of comparisons. Furthermore, the fact of Group 2's greater preciseness and uniformity as compared with the narrow-interval training groups on the narrow-interval test task prompts the suggestion that ease of discrimination appears not to be dictated solely by the size of the interstimulus interval presented to the subject but to be determined also by the extent to which his discriminatory mechanisms have been developed at the time of test. Indeed, it is not so much a matter of establishing the superiority of wide versus narrow training intervals as it is of determining the optimally discriminable interval relative to a given level of development. The present results on interval size parallel those of Lawrence (1952) reported while this experiment was in progress. The fact that his results were obtained with white rats rather than human subjects and a visual-intensity discrimination rather than a weight-discrimination task suggests that our preciseness data are independent of the present particular conditions of experiment. It also suggests the possible fruitfulness of further exploration of the relation between ease of discrimination (interval size) and transfer efficiency and its explanation both as a point of reorientation in learning theory and as a basis of more effective teaching technique.

(ii) *Subsidiary comment*

The present accuracy data have implications for Helson's (1948) and Johnson's (1944) studies of the position at which the subject 'zeros in' on the stimulus scale. Their results have shown that this statistic does not always stand in constant relation to the objective standard, but may, instead, vary with changes in the amount and magnitude of previous stimulation and with present properties of the context or ground. Our accuracy scores, meanwhile, do not conform to what might be expected from a consideration of their findings. If the reference standard (adaptation level) represents an average of all instances of stimulation, then the mean accuracy scores for Groups 2, 3 and 5 should be equal but

of greater magnitude than those of Groups 1 and 4. The failure of this prediction to be confirmed may be explained in several ways without invalidating adaptation level theory. Perhaps the most likely explanation is to be found in the temporal distribution of weight experience in the training groups. Helson's and Johnson's reports apparently are based on judgements made during a single experimental session. Our training and test sessions were separated by approximately 24 hr. The results of Group 4 may be accounted for if we assume that more similar (adjacent) stimuli exert an influence upon the weighted average of experience smaller in magnitude than that of the less similar (assuming both categories to occur with equal frequency) but one of longer duration. If the Helson and Johnson shifts are only temporary effects, then the degree of permanence with which they may be induced must be further studied, as must the trend toward high accuracy in the subjects of this experiment.

VI. SUMMARY

The relation of breadth of past experience to effectiveness of generalization was studied in the context of the classical weight-lifting experiment. Groups were trained about a standard of 300 g. with a wide-range, wide-interweight-interval series (Group 2), a narrow-range, narrow-interval series (Group 3), and a wide-range, narrow-interval series (Group 5). Another group was given a narrow-range, narrow-interval series about a 220 g. standard (Group 4) and still another (Group 1) made judgements only with the intermediate-range, narrow-interval *test* series about a standard of 100 g. Groups 2, 3 and 5 did not differ reliably from perfect accuracy in test performance. Group 4, however, displayed a small but significant positive constant error. Group 2 was, meanwhile, significantly more precise than the other groups, which did not differ from each other in preciseness. The greatest intra-group uniformity was found in Group 2, the least in Group 3. Superiority in test performance was thus found to be associated with interval size rather than breadth of training or greater adjacency of training and test weights on the stimulus continuum.

REFERENCES

AMES, A. (1946). *Some demonstrations concerned with the origin and nature of sensations.* A laboratory manual (preliminary draft).

BRUNSWIK, E. (1939). Probability as a determiner of rat behaviour. *J. Exp. Psychol.* XXV, 175–97.

CANTRIL, H. (1950). *The Why of Man's Experience.* New York: Macmillan.

CHAPANIS, A., GARNER, W. R. & MORGAN, C. T. (1949). *Applied Experimental Psychology,* ch. 5. New York: Wiley.

HARLOW, H. F. (1949). The formation of learning sets. *Psychol. Rev.* LVI, 51–65.

HARLOW, H. F. (1951). Thinking. In H. Helson (Ed.), *Theoretical Foundations of Psychology,* ch. 10. New York: van Nostrand.

HEBB, D. O. (1949). *Organization of Behaviour.* New York: Wiley.

HELSON, H. (1948). Adaptation-level as a basis for a quantitative theory of frames of reference. *Psychol. Rev.* LV, 297–313.

HULL, C. L. (1943). *Principles of Behaviour,* ch. 12. New York: Appleton Century.

JOHNSON, D. M. (1944). Generalization of a scale by the averaging of practice effects. *J. Exp. Psychol.* XXXIV, 425–36.

KÖHLER, W. & VON RESTORFF, H. (1933). Analyse von Vorgängen in Spurenfeld. *Psychol. Forsch.* XVIII, 299–342.

LAWRENCE, D. H. (1952). The transfer of a discrimination along a continuum. *J. Comp. Physiol. Psychol.* XLV, 511–16.

RIESEN, A. H. (1950). Arrested vision. *Sci. Amer.* CLXXXIII, 16–19.

WOODWORTH, R. S. (1938). *Experimental Psychology.* ch. 17. New York: Holt.

(*Manuscript received* 23 *February* 1954)

MEASUREMENT AND MATHEMATICS IN PSYCHOLOGY[1]

By A. S. C. EHRENBERG

Institute of Psychiatry (Maudsley Hospital), London, S.E.5

I. *Introduction* (pp. 20–21). II. *The spurious distinction between quality and quantity* (pp. 21–23). III. *The irrelevance of additive units* (pp. 23–25). IV. *The use of mathematics* (pp. 25–26). V. *The standardization and comparability of measurements* (pp. 26–28). VI. *Summary* (p. 28). *References* (p. 29).

1. Introduction

Psychologists tend to feel uncertain and apologetic about the numerical methods which they use in describing their observations. Mental test scores, ratings, sensory assessments and the like are often said not to be 'measurements' at all, and the validity of their mathematical analysis is doubted. Criticisms tend to be rather vague—some logical or even metaphysical attributes of 'measurement' are apparently thought to be lacking. More specific charges refer to the 'arbitrary' nature of the scoring-units, and to their alleged lack of 'equality' and 'additivity'. The consequences are implied to be dire, even if they are left obscure.

Criticisms arise because the measurements of physics, still called the 'science of measurement' even nowadays, are taken to be the necessary model of *all* measurement. But our familiarity with the conventional 'dimensions' and 'units' of physics makes us clothe them with a reality going far beyond the demonstrable empirical facts. Historically it would be inept, and factually wrong, to regard the degree scales of temperature, for instance, as derived from our present-day knowledge of 'absolute temperatures', 'equal amounts of work done per degree rise of temperature', and so on. Temperature scales were set up by the arbitrary, or better perhaps, the deliberate, subdivision of the change in something as 'far-fetched' as the length of a column of mercury; subsequent formulations have only led to equally conventional re-definitions of the scale in terms of changes in platinum resistances, thermo-couple e.m.f's, black-body radiations, and so on.

The importance of any 'theory of measurement' considerations may perhaps best be judged from physics. Here the physicist himself is generally said to have no knowledge of the 'theory of measurement', nor does he seem to have much use for it. Indeed, there is no accepted theory of measurement. Typically, the British Association Committee (1938, 1940) 'On the possibility of quantifying sensory events' could not agree on the meaning of measurement, and this is still the common state of affairs (cf. Bartlett, 1950). The various attempts which are made to develop some such theory appear doomed, like other contributions to the philosophy of science, to be ignored by the practising scientist. Recent papers on measurement in the psychological literature (e.g. Gulliksen, 1946; Coombs, 1950, 1951, Coombs, Raiffa & Thrall, 1954; Burke, 1953) lean on the better-known writings of the physicist Campbell (1928) and the psychologist Stevens (1951) which will be discussed below. It was also to these two authors that reference was made in discussions of measurement theory at the recent symposium on Flavour Assessment (1953). There the usual

[1] This article is based on a paper read at the British Psychological Society's Annual Conference, Nottingham, April 1953, and arises partly out of work carried out on behalf of the Department of Scientific and Industrial Research (Food Investigation Organization).

attitude to non-physical measurement was made explicit by Dr J. O. Irwin (1953) who, on the one hand, said that the derivation and use of certain scoring scales seemed to him to be 'thoroughly scientific' and, on the other hand, stressed the doubts that might be felt in treating such scales as if they had an 'underlying metric'. (Compare also Bartlett's comment on the earlier symposium (1950)—'When Theory whispers low "you can't", Practice replies "I must".') At the flavour symposium, Dr J. M. Shewan and the present writer (Ehrenberg & Shewan, 1953) tried to bring the matter to a head by suggesting that 'if any attention is to be paid to those people who say that one cannot "measure" in the present context for some reason or other, they should be asked to state specifically what they mean, in terms that are capable of being tested empirically'. There seems to be no instance of an actual mathematical analysis of numerical assessments which has been criticized purely on grounds of 'theory of measurement' criteria.

Nothing definite, then, seems to be known which would stop the empirical scientist from describing his observations in any language which he finds useful; he merely has to ensure that his definitions are not self-contradictory. A not uncommon fallacy here was exemplified by Prof. W. D. Wright who at the symposium on 'Subjective Judgments' already referred to (Bartlett, 1950) 'asserted categorically that sensation magnitudes cannot be measured, for there is no yardstick to be laid alongside of them'. The initial problem is how to define a yardstick for the first time, not how to measure along some given yardstick.

A definition of 'measurement' may therefore be something like 'the assignment of numbers to observable phenomena'. If the phenomena are consistently classified, some assignments of numbers will be more useful than others, but none will be invalid. Most formal definitions of measurement begin by some such statement. But the definitions are not left at that, for the assignment of *any* numerical symbol to *any* observable phenomenon is felt to make 'measurement' too arbitrary. Consequently, various attempts are made to restrict the meaning of measurement, usually following one or more of three lines of approach. One is to argue that only certain observable phenomena (e.g. 'quantities') are measurable, whereas others (e.g. 'qualities') are not; the second is to require that measurements should be 'physically additive'; the third is to attack the validity not of the supposedly more 'arbitrary' kinds of numerical labelling, but of the ensuing mathematics.

In the following discussion, these attempts to restrict the meaning of measurement will be queried.

II. The spurious distinction between quality and quantity

It is customary in discussions of measurement to differentiate between 'qualities' and 'quantities'. If these terms are supposed to distinguish observable phenomena which have not (yet) been numerically described from those which have, they are acceptable tautologies for the *achievement* of measurement. If, on the other hand, the distinction is taken to be an *a priori* one—e.g. that qualities cannot be measured because they are qualities—it of course begs the question.

The most extended, and also traditional, discussion of this topic in the more scientific (as opposed to philosophical) literature seems to be that of Campbell (1928, and elsewhere). Campbell says that 'measurable properties . . . are those things about which certain laws are true, the laws that remove the arbitrariness from the assignment of numerals and'—

here unfortunately he becomes obscure—'enable us truly to measure'. However, Campbell has to add later to this definition of measurement that 'even true measurement... is not completely free from an arbitrary element', and that in any case these 'laws of measurement' apply only to his so-called 'fundamental' magnitudes, and not to the apparently much larger class of 'derived' magnitudes. At the same time, he tells us that 'no accurate and important distinction can be drawn between the magnitudes that are measured by the fundamental method and those that are measured by the derived'. Indeed, Campbell's approach, not unlike that of other writers in this field, amounts to distinguishing not only between concepts such as qualities, quantities, magnitudes, properties, etc., generally, but also between fundamental and derived properties, measurable and unmeasurable qualities, *A*- and *B*-magnitudes, primary and secondary magnitudes and so on, without however showing that any of these classifications are logically watertight or of much demonstrable use.

If the attempts of Campbell and other such writers to reduce the arbitrary element in measurement are strangely unsuccessful, there still remains the fact that the physical scales of measurement do appear to us so very real and the converse of arbitrary. Their actual arbitrariness is, however, only hidden by long-standing conventions. Thus, Campbell's first 'law of measurement' was 'that the measurable property must have a definite order'; here the arbitrary choice of an order by some original scale-constructor is hidden by the idea that observable phenomena could 'have' an order in some absolute sense.

This matter of accepting the properties of physical measurement as if they were logical or empirical necessities should perhaps be briefly analysed here. In his recent book, *Measurement of Mind and Matter*, Scott-Blair (1949) adopts a common point of view when he says that it is clear that three dimensions are required to describe space. However, this is not only not clear, but demonstrably false. For example, all observable points, being rational, are enumerable, i.e. they can be described in *one* dimension.

Scott-Blair considers the description of houses in a symmetrical American town, with streets at right angles. Supposing, for simplicity, that there were seven streets in each direction (and that the ground is flat), the houses (or blocks) would be arranged something like the squares on a chess board. They could then be described in the usual, two-dimensional, way by pairs of co-ordinates such as:

1, 1	1, 2	1, 3	1, 4	1, 5	1, 6	1, 7	1, 8
2, 1	2, 2	2, 3	2, 4	.	.	.	.
3, 1	3, 2	3, 3	.	.			
.	.	.	.				
.	.	.					

But one-dimensional descriptions could be made up, for example as follows:

1	2	3	4	5	6	7	8
9	10	11	12	·	·	·	·
17	18	19	.	.			
.	.	.	.				
.	.	.					

With this latter map one could still find one's way around, perform calculations on distances, and so on. The effective difference between the two kinds of mapping is in the

'principles of closeness' (as I shall call them) which are involved. Thus, in the first, any points which are 'close together' in our usually adopted sense of distance (e.g. of putting one foot in front of the other) have similar co-ordinates, in marked contrast to the one-dimensional mapping.

Now it may be true that in general practice we find those kinds of space-measurements which obey the first principle particularly useful and simple. However, in describing the location of houses in a normal town, for example, a different principle of closeness can be more advantageous, and is in fact generally used, namely that all houses facing a given street should be considered 'close together' and be given similar descriptions, with the result that there are many dimensions (street names, and numbers) and that houses which are quite close together in the usual sense (e.g. having adjoining back-gardens) may have *all* their co-ordinates quite different.

Now it seems also to be true that if we adopt the *first* principle of closeness, two dimensions are required for an appropriate space-description here, or three dimensions in general. However, as we have just seen, this principle is not a *necessary* one, and consequently the use of three dimensions for space-measurement is not necessary. There may be something deeper, i.e. less 'conventional', in the three dimensions of space (although personally I doubt it). But before the measurement of psychological phenomena is dismissed for its arbitrary conventions, some direct evidence of the lack of comparable conventions in physical measurement would seem to be required.

Without attempting to go into much further detail, there seems to be just one other aspect of 'principles of closeness' which helps to explain the apparent difference between the status of physical and that of most non-physical measurements. The various steps made in defining a system of measurement, namely the choice of the number of dimensions, the choice of ordered sequences within each dimension, and the choice of unit steps, may each be said to derive from a 'principle of closeness'. Now in non-physical scaling, *different* principles of closeness tend to be required for these various steps; in contrast, the usual physical scales all seem to derive from a single principle (namely that of 'physical additivity'). It seems to be this derivation from a *single* principle, however arbitrary it may or may not be in itself, which has given physical scales of measurement their almost overpowering elegance.

III. The irrelevance of additive units

The belief that measurements are 'true' or 'valid' only if the units of measurement are *additive* is very common, although there seems to be little foundation for it. Additivity here seems to mean that the *arithmetical* addition of the measurements in question must represent, or be capable of representing, certain processes of *physical* addition for the observable phenomena, or vice versa.

Campbell's writings are again usually referred to here for a formal statement of the alleged need for additivity. He embodies the requirement in the second of his 'laws of measurement'. But apart from the general lack of cogency of these laws which has already been noted above, Campbell says (1928, p. 19) of this particular law that the process of constructing a scale ' could be carried out, even if that law were not true'. Admittedly, he adds that the *usefulness* of the process depends on the truth of the law; but that is at best a much weaker proposition—and is certainly open to question—whilst pragmatically interpreted it even contradicts the assertion that the law need not in fact be true.

In general, it is surprising that it should be the issue of 'additive' or 'equal' units which most befogs the theory of measurement. Consider some set of observable phenomena already put into some ordered sequence, A, B, C, D, etc. Then it matters little whether numbers like 1, 2, 3, 4, etc. are attached to them (making each interval A–B, B–C, etc., numerically equal) or numbers such as their squares 1, 4, 9, 16, etc. (making the intervals A–B, B–C, etc., increasingly large, numerically), and so on. The numbers in the different sets are functionally related, i.e. logically they are identical. It follows that it is impossible for one set of units to be valid and another not.

Even in physics, the original choice of units is unimportant. Most physical scales of measurement (e.g. temperature, density, etc.) are, of course, in any case not directly 'additive', but even when they are (e.g. length, time, mass, etc.) the use of the measurements seldom depends on the physical additivity of the units. The distance d fallen in time t under constant acceleration g by the famous apple starting from rest, is now known to be given by the relation $d=\frac{1}{2}gt^2$. If we happened to be measuring time (as T) in non-additive log-units (which are so common in bio-physics for example) and also measuring distances in non-additive logarithmic units as D, the relation would be equally valid, being expressed as $D-2T=G$, a constant.

As for the *usefulness* of the additive units, it will be noted that the logarithmic law, being *linear*, is much simpler than the one expressed in the additive units, d and t. For example, the analysis of two sets of observations on a falling body such as:

$t=0{\cdot}5$	1	1·3	1·5
$d=4$	16	27	36

and

$t=0{\cdot}5$	9	13
$d=4$	1296	2704

would be somewhat difficult. On the other hand, measuring in logarithmic units, we could hardly help noticing that the difference between each pair of readings D and $2T$ was a constant, namely 1·2041. Equivalently, having hit on the dodge of transforming the time measurements on to a non-additive scale of the *square* of t, the simple proportionality of t^2 and d in each set of data, i.e.

$t^2=0{\cdot}25$	1	1·69	2·25
$d=4$	16	27	36

and

$t^2=0{\cdot}25$	81	169
$d=4$	1296	2704

would soon become apparent, and merely by comparing the two constants of proportionality then—16 in each case—it would be seen that the two sets of data do, in fact, obey the same law, $d=16t^2$.

The physically additive units of certain physical scales of measurement are therefore *not* always the 'simplest' to use and we seem to continue to use them (when we do) only for the sake of standardization and communicability, since they happen to be the traditional ones.

Turning now to the possible usefulness of additive units in psychology, we find that for most psychological phenomena (e.g. mental tests, subjective ratings, sensory assessments, etc.) the concept of some kind of 'physical' addition is largely irrelevant.

One could, if one wished, define a process of physical addition of intellectual functioning, for example, as the joint performance of two people on the given 'intelligence test'. But scale-construction which is merely based on additivity of *some* kind however far-fetched seems to be a waste of time. This is shown, for example, by the various attempts in psychophysics to derive sensory response scales with units 'equal' in terms of physical additions of the stimulus-variable, or by the modern work in the field of 'scale-analysis' (such as the derivation of metric scales from 'paired comparison' data, based on so-called 'assumptions' about so-called 'underlying variables' which are to behave additively): none of these scales with formally derived 'additive' units ever seems to be of much practical use.

The belief that additive units are neccessary for measurement may also be questioned on quite a different ground. It is that no reasons seem ever to have been suggested, let alone shown to be cogent, why an empirical correspondence with arithmetical addition, which is only one out of all the many mathematical operations which have been defined and studied, should be *the* criterion of 'measurability'.

IV. The use of mathematics

Instead of attempting to deny the attribute of 'measurement' to certain numerical assessments, some writers doubt the validity of most of their mathematical analysis. Stevens (1951, and earlier), for example, first adopts a very catholic definition of measurement, but considers scales of measurement as being of four different kinds. He classifies them, as either Nominal, Ordinal, Interval or Ratio scales and lists for each kind of scale certain mathematical operations which he calls 'permissible'. Stevens himself has difficulty in fitting something as crucial as *rankings* into the classification, and generally it seems unlikely that it would stand up to rigorous analysis. But the important fact to note here is the lack of evidence that the use of mathematics can be thus circumscribed.

Consider, for example, the Nominal scale, which is supposed to refer to numerical labelling without any (explicit?) choice of order or of units. Only the simplest operation may, according to Stevens, be performed on Nominal measurements, for example the counting of frequencies, but not addition, multiplication, division, etc. (The more complex functions used in applied mathematics are never explicitly 'permitted' by Stevens at all.) However, it has been found that the numbers listed in telephone directories—an instance, presumably, of Nominal measurement—cannot be used very adequately in making up tables of random numbers, and all the complicated mathematics of 'tests for randomness' had to be used to investigate the matter. The attempt to restrict mathematical analysis thus does not agree with practice.

The contrasting point of view to be urged here is that applied mathematics must do neither more nor less than describe relations between observed phenomena. For this the mode of definition of the measurements if irrelevant—for to 'measure' in Wright's sense of *using* a given yardstick, and not of *defining* one (see p. 21 above), is to adopt a routine process of describing certain observations in an agreed manner which we have merely learned to apply by rote. One could say that the validity of a mathematical analysis of measurements depends not on a correspondence of the mathematical operations with empirical *operations* such as physical addition, but on a much more general kind of correspondence, namely with empirical *relations*. It may be noted here that this principle

of 'empirical validation' explains why classifications such as Stevens's and also Campbell's tend to appeal to most of us at first sight, even though on further analysis they turn out to lack relevance to our practical problems. The point is that Stevens's 'permissible mathematics' is that particular mathematics which is already empirically validated by the operations used merely in *defining* the scale of measurement. For example, suppose that the I.Q. and the weight (in pounds) of a certain child are both 110 and those of another 130. Then for the I.Q.'s we do not directly know of any meaning for the particular calculation $110+130=240$, although we may *later* find such a calculation to be meaningful. But for the weights we know at once—because of the physically additive definition of the units—that the two children put into one pan of a balance would, in fact, be balanced by a weight of $110+130=240$ lb. in the other pan.

Now, to show that the empirical validation of applied mathematics can, in fact, be sufficient, an example should perhaps be given which covers also the arguments criticized in the two previous sections, namely that the measurement of 'qualities', and addition without 'physically additive' units, are both supposedly 'illogical'.

Consider what is perhaps the most maligned mathematical operation, namely averaging. The simplest non-trivial case occurs where the observations fall into only two classes, which may be denoted or 'measured' by the numbers 0 and 1. Adding all the given measurements together and dividing by their total number gives then an average which is a proper fraction. Although no corresponding observable phenomenon or operation has been defined, this number can, of course, be interpreted, being the proportion of 1's in the data; and together with the total number of measurements, it will describe the observations completely. The process of addition and averaging is therefore meaningful, even though the measurements 0 and 1 stand for 'chalk' and 'cheese', or 'dead' and 'alive', observable phenomena which are often thought incommensurable.

V. The standardization and comparability of measurements

When people feel that something or other 'cannot really be measured', they often mean that two observations which have been given the same label do not necessarily 'mean' the same thing. A good deal of confusion arises because the latter problem has nothing to do with the applicability of numerical methods. In investigating under what conditions and in what respects two observations which have been given the same label are in fact comparable, it is irrelevant whether the label is a numerical one (a 'measurement') or not.

The problem of standardizing a new measuring technique, which is of course a fundamental one, requires for its solution nothing but a direct experimental attack (cf. Ehrenberg & Shewan, 1953). Critics of psychological measurement, and also those psychologists who would have us appalled at the task confronting them, often seem to ignore that the present-day degree of standardization and objectivity achieved in describing *physical* phenomena has only been reached after a great deal of detailed experimental work. It may now be known that the reading on certain thermometers will be about 100° C. when water relatively free from impurities is boiled at a controlled pressure; but it was not always as 'easy' as that. The work of actually investigating the conditions and respects in which a new measuring technique does give comparable readings has two aspects which require discussion here.[1]

[1] The unfortunate habit in the mental testing field of devising a new test, administering it to some arbitrarily chosen group of subjects, calling these 'the standardization population', and then leaving it at that, does *not* seem to call for comment.

On the experimental side there is the need for an explicit striving for standardized and 'objective' observations, and on the theoretical side there is the need for methods of analysis which facilitate the comparison of different sets of data.

First, the striving for comparable observations. It often happens in psychological work that a person has to assess or describe something the like of which he has never seen before. He therefore has to use descriptive terms or scores on whose appropriateness he cannot properly have agreed with anybody else. Clearly, no attempt can be made in such cases to obtain 'comparable' or 'standardized' or 'objective' results. Many such assessments are of course *deliberately* subjective, for example when a subject is asked to rate some picture 'on a 5-point scale for beauty'. It is irrelevant to the validity of such an assessment whether or not anyone else will agree with it. But such deliberately subjective responses do not in fact amount to the observations which the psychologist would report. Thus a subjective assessment is not reported as 'such and such a picture has a score of 4 for beauty' since others need not obtain the same result; instead, the observation is described as 'so-and-so scored the picture 4' with the implication that other qualified observers should also be able to observe so-and-so scoring 4. It follows that the occurrence of deliberately subjective assessments is no obstacle towards establishing under what conditions different observers could, in fact, observe the same results.

However, investigators in psychology, psychiatry, and in related fields often miss the point—or so it would appear—and are content for their own observations to remain on a subjective level. Thus, in many rating or sensory assessment situations, in psychiatric diagnosis, and so on, investigators tend to say what they personally consider appropriate, true, or proper, instead of trying as a primary step to achieve general agreement; but there seems to be no way (in science) of *knowing* what is appropriate, true or proper, if no one will agree. Again, even where objective techniques of observation are the avowed aim, the degree of standardization achieved is seldom investigated, let alone under constant scrutiny; but objectivity, in order to exist, must be known to exist, and be readily demonstrable.

Secondly, the need for methods of analysis which allow all experimental results to be readily compared. Rather surprisingly, perhaps, the so-called 'psychometric' methods especially developed or adapted for the analysis of psychological data are of a kind which either hinder the comparison of different experiments, or make it altogether impossible. The reason is that psychometric theory and practice deal largely in what are called 'standardized' variables. For a given set of readings, a variable is 'standardized' in this special sense if the mean of the readings is subtracted from each individual one, and if each such difference is divided by the standard deviation of all the readings. This gives a new scale of measurement which has for the given set of readings a zero mean and unit standard deviation. These properties are of course quite specific to that single set of readings; in the next experiment a *new* standardization will be carried out, and so on. The standardized scale is therefore always the same (zero mean, unit standard deviation) whatever the observations may be; or, to put it another way, the scale (in its relation to the actual observations) will differ from one experiment to the next.

Most frequently, perhaps, 'standardized' variables are used mainly by default. They are implicit, for example, when the relation between two variables is described by a correlation coefficient as follows: the correlation between variables x and y is 0·63. This merely tells us that a positive relation exists, i.e. that relative high values of x tend in

some unspecified manner to go with relatively high values of y. But it does not tell us what the relation between the variables actually is: *How* does variable x vary as y varies? To be able to answer this question, we would of course need to know the means and the standard deviations of the variables as well, and it is these latter statistics which are often either ignored (by loosely talking about 'the correlation') or deliberately omitted (by explicitly working with 'standardized' variables). It follows that it is largely meaningless to compare the relationships between two variables in different sets of data by means of correlation coefficients (even when corrected for attenuation), i.e. to suppose that one has the same relationship merely because the correlation coefficients happen to be the same.

A more explicit use of 'standardized' variables occurs in what is called 'profile analysis'. Here the variables in any one study are supposed to be made 'comparable' by their standardization. But profiles from different studies are not comparable.

A more extreme case still occurs in factor analysis, where not only the *observed* variables are standardized in each analysis, but also those of the factor model itself. The complete absence in the literature[1] of valid techniques for comparing factors extracted from different analyses is therefore logically quite correct, even if not helpful.

The psychometrician's reason for using 'standardized' variables—when he makes it explicit—seems to be something like that 'he has no proper scales of measurement', or 'that he does not know what his measurements mean, anyway'. It seems certain that he can never overcome the first difficulty by using a different scale of measurement in every new experiment, or the second by making it impossible to compare the results of the different experiments. Surely the proper purpose both of numerical measurement and of applied mathematics must be to describe empirical observations in such a way as to facilitate their comparison with each other?

VI. Summary

Measurement may be said to be the assignment of any numbers to any observable phenomena. It cannot be restricted to any one kind of phenomena or scale. Thus 'qualities' differ from 'quantities' merely by being phenomena to which no numbers have as yet been assigned. Again, physically additive units of measurement, which for many kinds of observation are so obviously irrelevant, are also logically unnecessary; this is shown, for example, by the fact that the choice of arithmetical addition as the mathematical function involved is a purely conventional one. Applied mathematics can be used to simplify the description of any observations, and seems to need no other justification.

The doubtful validity or usefulness of many 'measurements' arises therefore not from anything in their numerical nature, but because the underlying observational techniques either are not standardized and objective, or have at least not been shown to be so. In psychology this often follows from the use of 'psychometric' methods of analysis based on so-called 'standardized' variables, which make it impossible to compare the results of different experiments.

[1] One of the few explicit attempts to fill this gap seems to be in Cattell's text-book (1952, chap. 17), but it is necessarily unsuccessful. More recently Barlow & Burt (1954), for example, tell us that 'the best measure of identity (of two factors) will be the correlation between the two sets of factor-measurements'. Even if we ignore the fundamental impossibility of obtaining any factor-scores (Ehrenberg, 1953, p. 140), this proposed criterion of course suffers from all the failings of correlational analysis referred to above. Suppose, for example, that a factor extracted in one study is completely 'identical' in this sense with a factor extracted in another study, then persons with identical test-scores in the two studies will not in general have the same factor-scores.

References

Barlow, J. A. & Burt, C. (1954). The identification of factors from different experiments. *Brit. J. Psychol. statist. Sect.* vii, 52–6.

Bartlett, F. C. (1950). Subjective judgements. *Nature, Lond.*, clxvi, 984–5.

British Association (1938). The possibility of quantitative estimates of sensory events. (Reports of special committee). *Advanc. Sci.* p. 227.

British Association (1940). The possibility of quantitative estimates of sensory events. (Reports of special committee). *Advanc. Sci.* p. 332.

Burke, C. J. (1953). Additive scales and statistics. *Psychol. Rev.* lx, 73–5.

Campbell, N. R. (1928). *Measurement and Calculation.* London: Longmans, Green and Co.

Cattell, R. B. (1952). *Factor Analysis.* New York: Harper.

Coombs, C. H. (1950). Psychological scaling without a unit of measurement. *Psychol. Rev.* lvii, 145–58.

Coombs, C. H. (1951). Mathematical models in psychological scaling. *J. Amer. Statist. Ass.* xlvi, 480–9.

Coombs, C. H., Raiffa, H. & Thrall, R. M. (1954). Some views on mathematical models and measurement theory. *Psychol. Rev.* lxi, 132–44.

Ehrenberg, A. S. C. (1953). Critical Review of Cattell's 'Factor Analysis'. *J. Genet. Psychol.* lxxxiii, 185–96.

Ehrenberg, A. S. C. & Shewan, J. M. (1953). The objective approach to sensory tests of food. *J. Sci. Fd Agric.* iv, 482–90.

Proceedings of symposium (1953). Flavour assessment. *Chem. & Ind.* pp. 562–5.

Gulliksen, H. (1946). Paired comparisons and the logic of measurement. *Psychol. Rev.* liii, 199–213.

Irwin, J. O. (1953). Flavour assessment: A biometrician's viewpoint. *Chem. & Ind.* pp. 1348–52.

Scott-Blair, G. W. (1949). *Measurements of Mind and Matter.* London: Dobson.

Stevens, S. S. (1951). Mathematics, measurement and psychophysics. In *Handbook of Experimental Psychology* (ed. S. S. Stevens). New York: Wiley.

(*Manuscript received* 19 *February* 1954)

A THEORY OF SHAPE RECOGNITION

I. *Introduction* (p. 30). II. *The Problem* (pp. 30–31). III. *The System* (pp. 31–36). IV. *Summary* (pp. 36–37). *References* (p. 37).

By J. A. DEUTSCH

Institute of Experimental Psychology, University of Oxford

I. Introduction

The purpose of this paper is twofold. First, it is intended to show by another practical example the fruitfulness of a view of psychological theory put forward in another paper (Deutsch, 1953). In this the author proposed a theory of motivation and learning which was of a particular logical nature. It was argued that the task of a theory in psychology is to infer the system or mechanism which is producing the phenomena of behaviour without necessarily giving the embodiment of such a system. That such a policy was practically feasible at this stage was shown by the construction of a system whose properties tallied closely with those of a motivated learning organism. Further, to show that this theory was really the description of a mechanism in general terms and not a re-description of the behaviour of an animal in highfalutin language a machine was actually built embodying the system. It is intended here to give another example of such a theory and to show how this approach is of use in the study of perception.

The second object is partially to fill what many felt to be a gap in the previous theory. The system described in the previous paper dealt with and acted on information which had already been digested. The process of classifying environmental stimulation was taken for granted. In what follows an attempt is made to elaborate a mechanism of shape perception which is complementary to the behaviour system. Though there have been previous attempts in this direction, none of the models suggested has had properties which tallied sufficiently with the known peculiarities of shape recognition in animals. Nor was it possible to put them to test as they made no particular further prediction about behaviour. Accordingly they have more the status of inventions than of explanatory theories.

II. The problem

What are these known peculiarities which any theory concerning the subject must explain?

There are six main facts:

(*a*) Animals can recognize shapes independent of their location in the visual field. It is not necessary to fixate the centre of a figure in order to recognize it nor need the eyes be moved around the contours of a figure.

(*b*) Recognition can be effected independently of the angle of inclination of a figure in the visual field. (By this is not meant tilt of the figure in depth such as occurs in shape constancy experiments, but the tilt of an image in two dimensions.)

(*c*) The size of a figure does not interfere with the identity of its shape. This of course does not hold at the extremes of size for reasons which seem sufficiently obvious.

(*d*) Mirror images appear alike. Both rats and human beings tend to confuse these. This would tend to rule out any 'template' theories of shape recognition. According to

these a contour is rotated in two dimensions until it coincides with one of the many patterns already laid down. But such a superimposition cannot take place in the case of mirror images and no room is left for this particular type of confusion.

(*e*) Visually primitive organisms such as the rat and the octopus find it hard (perhaps impossible) to distinguish between squares and circles. This does not seem to be a limitation imposed by the peripheral characteristics of the optical systems. It appears to be a more central defect as these organisms can distinguish shapes which are far more alike geometrically. This type of evidence tends to cast doubt on theories which base themselves on the angular properties of figures. To quote Lashley (1938): 'The pair of figures giving greatest difficulty of discrimination where the element of acuity might play a part are the square and the circle. In several experiments these have required from 60 to 300 trials for learning and only one-fourth of the animals have shown any improvement above chance. Yet acuity is certainly not the chief limiting factor in this case.'

(*f*) These abilities, which appear to be mediated by the striate cortex, survive the removal of the major part of it. It is therefore reasonable to suppose that this ability to disentangle shape is common to all parts of the striate area and that one part of the striate area is not essential in helping the next one to operate.

This consideration would tend to rule out notions based on a scanning process. It is difficult to see how a regular scan could be maintained in the presence of extensive damage. Further, any system which requires the fixation of the centre of a figure so that it coincides with the centre of the visual field must also be ruled out. The ability is maintained even when there are extensive scotomata of central origin.

In what follows a system is proposed which will abstract from and yet satisfy the above specifications. It will also lead to fresh predictions about the discriminability of certain figures. It is therefore testable.

It is not suggested that it is the only system concerned in recognition. There must be other systems devoted to the recognition of brightness, location and size. This makes many experiments on pattern discrimination difficult to assess from the point of view of evidence for the mechanism of shape discrimination pure and simple.

III. The system

Let us assume that there is an array of units (or cells) arranged in two dimensions, each unit having many neighbours. This plane composed of cells has messages arriving on it from light receptors (or the retina). Each group of cells is joined to a particular retinal element in such a way that neighbouring retinal elements also excite neighbouring groups of cells on this two-dimensional array. But it is assumed that only a particular source of excitation is passed on from the retina. This is excitation from contours, the regions of the sharpest change.

Thus when a ball is focused on the retina the contour of this ball is projected on this two-dimensional array. Excitation in the form of a circle will appear on it. Thus we have the first rule.

1. Each unit on the two-dimensional array can be excited by a contour falling on the region of the retina to which it is joined.

2. When such excitation from the retina arrives each unit will pass on a pulse down what will be called a final common cable. It will also excite its neighbour. This neighbour will pass it on in its turn. Therefore when a contour is projected on the two-dimensional

array two things will occur: a message consisting of one pulse will be passed down the final common cable by each cell or unit on which that contour lies. Thus a measure of the total number of cells stimulated will be passed down.

3. Second, the contour will excite all the cells which lie next to it on the two-dimensional array. These will pass the excitation on to their neighbours but not down the final common cable. The assumption is made that a cell will pass on its excitation at right angles to the contour of which it happens to be a component. Thus for instance the excitation produced by a straight line will be a line equal and parallel to it and gradually moving away. In the case of a curve, for instance an arc of a circle, it will be also an arc of a circle moving away 'parallel' to it, but gradually decreasing as it draws away, coming to a point and then fanning out again, this time reversed (since all the advancing points of excitation move at the same speed from the point of their origin).

4. As such lateral excitation from a point in a contour advances another message will be sent down the final common cable as soon as it coincides with another point in a contour imposed on the two-dimensional array.

This message down the final common cable will at each moment give a measure of the number of points thus brought into coincidence. For example, if the figure is a rectangle, the first set of pulses sent down the final common cable will be a measure of the total number of cells excited lying on the contour (rules 1 and 2). The second set of pulses down the final common cable will be sent when the excitation travelling in a parallel line (rule 3) coincides with the two longer sides.

As these two sides are parallel to each other the excitation and the contours will coincide simultaneously and send a number of pulses proportionate to the length of these sides down the final common path all at the same instant of time. Some time later the disturbance from the two shorter lines will reach each of these lines and another message will be sent down the final common cable, indicating that a smaller number of cells have been fired off this time but also all simultaneously. Thus the message sent down the final common cable for any rectangle will consist of three sharp volleys, the proportions of the last two being governed by the ratio of the longer to the shorter sides. The message generated by a square will be two sharp volleys, both equal in magnitude. As the two pairs of sides are equidistant the advancing disturbances will hit the sides simultaneously. As they are parallel, the disturbance generated and the contours will overlap all along their length at the same instant of time, causing the cells representing the contour to fire down the final common path together.

Oddly enough a circle too would generate two sharp volleys, both equal in magnitude. All the pairs of points lying opposite each other are equidistant. Therefore the disturbance travelling across would reach all the opposite parts of the contour simultaneously—a sharp pulse equal to the first, when the circular contour first arrived on the two-dimensional array. Thus, if we agree that this is the number and relative size of the pulses that is made use of at the next stage in primitive organisms in distinguishing shapes, then circles and squares will be treated as identical. Thus the theory so far would predict the fifth finding (fact *e*) about the confusion between squares and circles. There is another group of predictions which appear *prima facie* unlikely and which have not been tested. This is that a pentagon and all odd-sided regular polygons, till there is a breakdown of acuity (central and peripheral) ought to be discriminable from a circle by these organisms where a square, a hexagon, etc., would not be.

To make the logic of this deduction clear let us consider an equilateral triangle, the simplest of the regular polygons without parallel sides. As soon as this is projected on the two-dimensional array a message will be sent down the final common cable giving a measure of the total length of contour. Immediately after this the three sides will propagate lines of disturbance parallel to themselves. As these lines of disturbance meet the contours of the triangle, there will be another message sent down the final common cable. But this time the cells on the contour will not fire off all together as the disturbance will pass across them obliquely. Their firing will be staggered according to the angle at which the disturbance and the contour interact. Thus in the case of the equilateral triangle, there will be a large initial spike or sharp volley followed immediately by a much smaller continuous discharge which will suddenly cease.

In the case of the pentagon then there will be a large and sudden first spike (as with all figures) followed by a gap (before the disturbances make contact with the opposite sides). There will then be a message of a much smaller magnitude and spread out in time, following it as the disturbance meets the opposite contours obliquely. The pentagon then, given this system of encoding and decoding ought to be quite easily discriminable from both a circle and a square, whilst these latter appear the same.

In spite of this interesting confusion it would seem that this system would derive fairly distinctive messages from most shapes which cannot be superimposed. There are two exceptions to this which are again important. This is the case of mirror images (fact *d*), another interesting property which seems accounted for. The same message will be sent down the final common cable as long as the relations of angle to length of side remain the same internal to the figure.

The second relates to figures with highly complex and irregular contours and internal lines. Unless there is some regularity of geometrical arrangement within complicated figures, the messages down the final common cable generated by these will tend to approximate with growing complexity. There would be a very large initial spike followed by a continuous signal as the many disturbances set up by contours hit other contours continuously. Any feature of equal or parallel spacing, which we normally call regularity, would set up messages which were synchronous and thus additive or sharp, giving rise to distinctive peaks in the message. Absence of these would lead to the opposite. This tallies with Lashley's (1938) experimental findings. 'The indications are that a somewhat larger number of animals fail to discriminate irregular than regular geometrical figures, but those which do discriminate require no longer to learn the one than the other type. There is evidence that those animals which learn do so by isolating some simpler cue or part figure.'

The system outlined would also display the properties of sending the same message down the final common cable whatever the inclination or position of the figure on the two-dimensional array (facts $a+b$). The message generated depends entirely on the properties of the figure itself and not on its relation with some external axes.

So far only one fact of the original six is left unexplained. This is the animal's ability to recognize shape independent of size (fact *c*). The encoding two-dimensional array will of course pass on information derived from differences in size though it discards differences in position. However, the same proportions between the components of the message passing down the final common cable will be preserved provided that the shape alone is constant. The problem of reducing messages of identical proportions to the same message

or having a detector which is only sensitive to the proportions independent of size must therefore be separately considered if we adopt this scheme. A way of doing this can be envisaged, especially as the first part of the message in the system outlined above could provide an indication of the scale of the figure. The difficulty is to decide between possible alternatives as there are hardly any clues. There is a constant relationship between the initial pulse giving the length of the contour, and the size of the rest of the following message for each particular figure. This relation may be used to reduce the message for each figure to the same size, whatever the original size of this figure. If the size of the second part of the message is always divided by the size of the first, the crests of the message for a given shape will not only be of the same proportions but also of the same absolute height. (The problem of the recognition of a figure now becomes merely the problem of recognizing pulses of identical height.)

How could such a scheme be worked? For instance, the measure of the total number of cells involved signalling the length of the contour could pass on ahead. It could then act on a section of the final common cable through which the whole message would soon pass on its way to the final detector units. It would throw this section of the final common cable into a state of refractoriness proportionate to the magnitude of this measure. It would thus alter its threshold of response to the following message to differing degrees. Only a very small proportion of a message following a large pulse would survive. A large proportion of the measure after a small pulse had set the 'resistance' would pass. Thus if both these messages were derived from the same shape, let us say at different distances, they would both give rise to a message the height of whose pulses was physically identical. Hence they would pass to the final stage whither we shall not follow them. We should pass too far into the realms of speculation to give an answer to what no longer seems a problem.

So far a mechanism has been suggested to reduce a purely relational entity, a shape regardless of its location, size and inclination, to a unique set of physical events. This mechanism also shares some other rather peculiar properties with an animal's shape-recognition mechanism. It is therefore supposed that such a mechanism is in fact that used by an animal.

This mechanism has other properties which follow as predictions.

1. It should prove easier to discriminate between a square and a rectangle than between a square and a rectangle both with one side missing.

2. There should be an almost perfect transfer from an equilateral triangle to another with one side missing. This agrees with Lashley's findings, but this type of transfer should be worse in the case of other triangles, where the removal of a side should make a great difference. For instance the removal of the hypotenuse in the case of a right-angled triangle should militate more strongly against transfer.

3. Shallow curves should be undistinguishable from straight lines in the absence of other contours in the vicinity. (It may be that the factor of a change of direction in the case of the curve may provide a cue here.)

4. If shallow curves and straight lines are indistinguishable, their addition to otherwise identical figures should make these discriminable. For instance, the substitution of a curved side in a rectangle, should differentiate it from the normal rectangle. This effect should be smaller for most triangles.

5. A 'mixture' of a square and a circle should be distinct from these even though they are not distinct from each other. In fact transfer to a rectangle ought to be easier in some

circumstances, for instance, where two arcs of a circle complete a square with two opposite sides missing.

6. A scotoma of central origin when falling wholly inside a figure, where there are no contours in the vicinity, should prevent identification of the shape.

7. The prediction concerning regular polygons with an odd number of sides has already been made. It will be noted that all these predictions are independent in the sense that it would be difficult to subsume any two under a single generalization. If to the writer's greatest surprise some of these predictions receive experimental confirmation, it may be worth drawing a few more deductions from the theory.

There are, however, some difficulties.

(i) Such a mechanism could only handle one figure at a time. Consider the effect on the message down the final common cable of a square and a triangle in two different parts of the field. The two messages would be inextricably mixed up or superimposed as they would be sent down at one and the same time. They should therefore mask each other to a certain extent. But they are in fact recognized easily and separately except in one interesting condition. This is when the contours actually touch or cross. It is therefore suggested that figures are passed to the two-dimensional array on the following principle. Assume that the points in the visual field are arranged in an arbitrary order of precedence. The stimulated point which is highest in order will inhibit all the others, preventing their passage until another signal (such as might arise from the final identification apparatus) is sent up. Then the next point in order of precedence will be dealt with. Now consider the case where such points have neighbours. Assume that the effects of this inhibition are overcome by a supervening excitation passed on from this point which is inhibiting all others to those of its neighbours (if any) that were also excited, and passed on by these. (It is assumed that interrupted contours are completed.) In this way a whole contour is passed on (and other contours are kept in 'cold storage') until it is dealt with. If this was the case, masking of one figure by another would occur as soon as their outlines touched. They would now be dealt with as one complex figure. Something of the sort is borne out by everyday experience and Gottschaldt's experiment (1926). In any case this is merely to show that this property of being able to handle only one shape at a time can be circumvented by an additional hypothesis.

(ii) Higher forms can distinguish squares from circles. There is no evidence that human beings have any difficulty at all in this. It would be more plausible to propose a mechanism which would preserve a phylogenetic continuity. The present scheme cannot be made to distinguish between squares and circles as it stands. But there can be quite a simple 'evolutionary' development of it to work with greater precision and to make this distinction. The encoding part of the mechanism can be kept the same and the same message sent down the final common cable. But the message must be sent to a detecting mechanism which is sensitive not merely to absolute differences in the height of the successive pulses. A mechanism is incorporated which is sensitive to the relation between the height of the pulses and their spacing in time. These will of course depend on each other for each particular form or figure. But their proportions may vary from figure to figure, even though the proportions of the successive pulses generated by these figures may be identical. This is in fact so in the case of the square and the circle. A circle's boundary is shorter than that of a square when the distance to be travelled by the laterally spreading disturbance from contour to contour is the same. Therefore, though the two pulses sent down in each

case are of the same height, there will be a greater distance between the two pulses even though they are of the same height. The analysis of the encoded information by two different means would probably improve the accuracy of recognition even where no radical confusion was possible. Such a possibility of improving the system by further development without any basic change perhaps gives the whole suggestion an enhanced biological plausibility.

(iii) It will doubtless be objected that nothing has been done to explain the lateral spread with the convenient properties postulated for it. No modus operandi has been suggested. This is admittedly a weakness of the hypothesis logically speaking. We ought in principle to state what family of machine performs this task, what interrelationships we postulate among our elements. On the other hand from the point of view of practical science to suggest possible networks of some complexity merely because they are required theoretically would be somewhat foolish, especially as there is nothing to distinguish between rival possibilities. It is really only necessary to ask whether it is plausible to assign such a property to the nervous system. Has anything like it in fact been found?

Motokawa claims to have found evidence for a laterally spreading effect on the retina, which he calls 'retinal induction'. 'For all these phenomena to take place it would be necessary that some process is propagated in a successive manner from one part to the next contiguous part.... In other words, the front of propagated induction may be supposed to proceed with a certain velocity through the retina, leaving a field of retinal induction behind it' (Motokawa, 1951). The processes of what Motokawa calls indirect induction show some of the features of the lateral spread which are postulated and which seem in many ways the least plausible physiologically. This is the ability for two types of indirect induction to cross over each other without interference with each other's spread. This property is clearly brought out in the phenomena of neutralization. Here the two processes generated by complementary colours invade each other's territory and neutralize each other. They do not form a contour as would happen if they nullified each other.

The fact that similar processes appear to take place on the retina would argue for the ease and simplicity with which the proposed system could be embodied in the central nervous system. Though no great weight is put on this analogy, it does show that the postulation of a process with somewhat similar properties in a structure of greater possible complexity is not too far fetched. It is true that what is so far known of this retinal induction indicates a greater crudity of operation than that which has been postulated for the central process. The further objection that such a central spreading process has not been suggested on anatomical grounds can be answered by saying that no-one suspected retinal induction on anatomical grounds either.

Of course until more work is done on retinal induction this possible analogy cannot be stressed too much. Perhaps the most one should say about it is that such a laterally propagated retinal process is being considered by some neurophysiologists.

IV. Summary

A theory of form recognition, complementary to a previously published theory of motivation and learning is put forward. A number of facts which any theory must account for, is given. These are ability to abstract form independent of place, inclination or size,

equivalence of shape of mirror images and of squares and circles in some species, and lastly the survival of these abilities following extensive lesions of the striate area.

A mechanism which displays the requisite properties is suggested. This mechanism is also seen to have further properties which ought to manifest themselves in behaviour if the appropriate tests are made.

Lastly various objections are discussed. Aspects of the mechanism are not entirely without neurophysiological plausibility.

References

Deutsch, J. A. (1953). A new type of behaviour theory. *Brit. J. Psychol.* xliv, 304–17.

Gottschaldt, (1926). Ueber den Einfluss der Erfahrung auf die Wahrnehmung von Figuren. *Psychol. Forsch.* viii, 261–317.

Lashley, K. S. (1938). The mechanism of vision. XV. Preliminary studies of the rat's capacity for detail vision. *J. gen. Psychol.* xviii, 123–93.

Motokawa (1951). Propagation of retinal induction. *J. Neurophysiol.* xiv, 339–51.

(*Manuscript received* 7 *May* 1954)

MOTOR AND MEMORY RESPONSES OF NEUROTICS AND NORMALS IN THE LURIA ASSOCIATION-MOTOR TECHNIQUE

By A. D. B. CLARKE

Senior Psychologist, The Manor Hospital, Epsom

I. *Introduction* (p. 38). II. *The population* (p. 38). III. *Procedure* (pp. 38–39). IV. *Results* (pp. 39–42). V. *Selective remembering* (pp. 42–43). VI. *Summary* (p. 43). *References* (p. 43).

I. Introduction

Luria's method (1932) is sufficiently well known to make a detailed explanation unnecessary; essentially it consists of a union between verbal and motor responses in the word-association process. According to Luria, by connecting the verbal response to the motor reaction of the hand, a system can be created capable of reflecting the whole dynamic character of the central associative process; this union gives to the investigator a method by which it is possible to estimate changes in an obscure process by examination of clearly defined and easily accessible motor reactions. Thus, every sharp fluctuation and every tendency to a speech response, every marked central affective disorganization, can be observed and analysed in a compound motor reaction.

Some interesting work has been reported on the various problems of the Luria technique during the last twenty or so years, and this has been reviewed in detail elsewhere (Clarke, 1950). Suffice it to say that these experiments have shown the essential correctness of Luria's thesis. For this reason, it was decided to incorporate an association-motor experiment in a research to determine the value of objective tests as measures of neuroticism.[1] Only the main positive results of the Luria experiment are reported here; for fuller details of methodology and procedure see Clarke (*loc. cit.*).

II. The population

The subjects consisted of forty-three neurotics who were all about to be discharged from the Army on the grounds of their psychiatric disabilities and forty-three 'normal' soldiers from an Army depot. The average ages of both groups were in the early twenties and, as will be seen later, the neurotics were significantly less intelligent than the controls; this latter was probably at least in part the result of a selection factor.

III. Procedure

A word-association list containing fifty words was drawn up, a few of these being neutral; the list was as follows:

1. Boat	11. Marriage	21. Prostitute	31. Bright	41. Germs
2. Ruler	12. Fire	22. Headache	32. Persecution	42. Father
3. Friends	13. Mother	23. Me	33. Late	43. Gravy
4. Tired	14. Nerves	24. Circle	34. Juice	44. Nice
5. Vomit	15. Naked	25. Drugs	35. Children	45. Frame
6. Reason	16. Ink	26. Dance	36. Fear	46. Leap
7. Intercourse	17. Inferiority	27. Castrate	37. Lamp	47. Colour
8. Home	18. Family	28. Wife	38. Coat	48. Venereal
9. Parade	19. Neutral	29. Interest	39. Kiss	49. Pain
10. Rain	20. Book	30. Murder	40. Grace	50. Flower

[1] This work formed part of a research programme at the Maudsley Hospital directed by Dr H. J. Eysenck and described by him in *The Scientific Study of Personality*, 1952, London, Routledge & Kegan Paul Ltd.

The apparatus consisted of an ink-pen continuous record kymograph, an electrical 'memory drum' which presented each word in turn to the subject through a small window at 10-sec. intervals, and a pneumatic linkage for the plate on which the right hand rested.

There was one addition to this test, aimed at casting some light on the process of 'selective remembering' in relation to the list of stimulus words. There have been several investigations in the past into the concept of repression, by means of word lists being shown to the subject, and subsequent attempt by him to recall their contents. As Zeller (1950) points out in a recent review of the whole field of experimental investigation of this concept, no really significant findings emerge. Some of the writers he quotes had assumed certain words to be 'pleasantly' or 'unpleasantly' toned, thus introducing an element of subjectivity. In order to find out whether words remembered in this way are those which are pleasantly toned, and those forgotten of unpleasant affective value, it is necessary to have some objective criterion for 'unpleasant affective value'. Thus, at the conclusion of the main test the subject, without previous warning, was asked to note down in a 3-min. period all the stimulus words that he could remember. It was known that this procedure would show certain words to be more frequently remembered than others. The question thus became: Would those words which produced the greatest motor disturbance on the test be those which were least frequently remembered by the subjects? An affirmative answer would support the concept of repression or 'active forgetting' in the test situation. In addition, a second question was asked, namely: Would differences in selective remembering be apparent in the two groups, i.e. would the neurotic tend to remember different words from the controls? A final point of interest was whether there would be a significant difference in the *number* of words recalled by the neurotic and control groups; it was predicted that this measure would be positively correlated with intelligence.

Before the test commenced the subject was told that this was an experiment to see how quickly he could react, and that as each word appeared he was to say the first word that came to mind, no matter what it was, at the same time pressing down the right hand plate and keeping his left hand passively resting on a lever. Sufficient practice was given to ensure that the subject was clear about what he had to do.

IV. Results

It was found that motor reaction times, failures to make motor responses, and small and large motor disturbances (measured as deviations from the straight line of the record) all showed significant point-biserial correlations with neuroticism. In addition, all these measures correlated significantly negatively with intelligence (Mill Hill Vocabulary Test). Thus, if the effect of intelligence on these scores was held constant by the part-correlation method, the correlation between the particular measure and the criterion (neuroticism) was somewhat reduced.

The best score for differentiating the two groups was found by the addition of the following items: (1) long motor reaction times, (2) right-hand 'small' disturbances, and (3) right-hand 'large' disturbances. The results gained from this composite score are shown in Table 1.

Correlation of this score with the Mill Hill Vocabulary Test gave an r of $-0{\cdot}590$. Using the motor disturbance score of 29 and above as an indicator of neuroticism, twenty-two subjects out of the eighty-six would have been misclassified (eight neurotics and fourteen control subjects).

If the effect of intelligence on this composite motor disturbance score is held constant by the part-correlation method (for this sample intelligence and neuroticism correlating to the extent of −0·320) then the correlation between the composite score and the criterion is reduced to 0·364.

The reader will have noted that all the aspects of motor disturbance on this test were negatively correlated with 'intelligence' as measured by the Mill Hill Vocabulary Test. Such a relationship seems not to have been noted in previous work, although Tendler (1945) reports it in regard to reaction times.

Table 1

	Neurotics	Controls
Mean	52·256	26·791
Variance	694·481	398·741
N	43	43

r_{pb} (neuroticism) = 0·483.
F ratio = 25·810, significant at the 0·001 level.

The Mill Hill Vocabulary Test is a test of verbal understanding; it might thus be possible that the neurotics, whose Mill Hill scores were significantly lower than those of the controls, showed disturbance as a correlate of lack of understanding. Three of the words from the stimulus list, 'Castrate', 'Inferiority' and 'Persecution', not infrequently elicited the response 'I don't know what it means'. Other words, such as 'Vomit' and 'Prostitute' were responded to in this way less frequently. One way of determining whether disturbance was partly a product of lack of understanding was to take a series of words which *a priori* and empirically would seem to have been understood by all subjects, which also showed differential disturbance on the motor level, and correlate the disturbance score for these words with the Mill Hill score to see whether a significant correlation also obtained. The following words were selected: 'Pain', 'Interest', 'Me', 'Late', 'Tired' and 'Colour'. On these words the point-biserial correlation of the composite disturbance score with neuroticism was 0·593. Correlation with the Mill Hill Test gave an *r* of −0·483, which should be compared with the correlation of the disturbance score on all the words of the list with the Mill Hill score: −0·590.

The fact that a significant correlation also obtains using stimulus words in which verbal understanding on the part of the subjects is likely to have been perfect, would seem to dispose of the hypothesis that the negative correlations with the verbal intelligence test were due to the difficulty of some of the words themselves. Previous investigators have viewed disturbance in association as being due primarily to affective interference, though it is true that the recent work of Albino (1948) leads him to suppose that: 'It is clear that the quantitative estimate of lability...from the graphic records shows that... (it) is not a simple quality, but is dependent on several factors.'

It will be remembered that Luria postulated a 'functional barrier' which by a regulating process kept conflict confined to cortical areas until adequate motor outlets were found. Thus the subject with the weak functional barrier showed greater motor disturbance on these tests, while the records of the subject with a strong functional barrier showed integration of response and little disturbance. Luria writes that 'in studying the genesis of the reaction movement, we see that the functional barrier does not exist in the early years of childhood but is elaborated rather late. Experimentation shows that it *appears*

about the time of the development in the child of the active organization of speech, and that it is absent in cases of serious mental retardation. All this connects the functional barrier with the higher psychological mechanisms, and the data of aphasia *place it in direct relation to speech and symbolic mechanisms*,[1] the falling out of which produces a diffused reaction...'.[2] In relation to the above, the present findings seem to support the concept of the functional barrier, and are quite in keeping with Luria's work. What is surprising, however, is that neither he nor later workers considered the natural corollary of the above views, namely that verbal intelligence would be negatively correlated with motor disturbance (as demonstrated in the present investigation), or alternatively, the greater the intelligence, the greater the tendency towards a 'strong' functional barrier, or vice versa. This has more than academic interest, since these previous workers have treated motor disturbance as being only the effect of emotional arousal, and have neglected this further aspect.

It will have been noted that the correlation of the disturbance score with the criterion was given in both 'corrected' and 'uncorrected' forms, the former giving the criterion correlation after the effect of intelligence had been held constant. This procedure assumes that neuroticism and intelligence are orthogonal, but numerous researches, like the present one, have reported small significant negative correlations between neuroticism and intelligence. In the present study this could have been, partly at any rate, the effect of sampling. The question thus becomes, therefore, which of the two values of criterion correlations represents the 'true' value? The present research has shown that the lower the intelligence, the greater the disturbance and therefore the weaker the functional barrier; it has also shown that the lower the intelligence the greater the neuroticism. In other words, a weaker functional barrier is associated with greater neuroticism and lower intelligence. These results, showing an interdependence of these factors, are consistent with the view that lower intelligence may constitute part of the criterion, neuroticism. If this is so, then the uncorrected criterion correlation represents the 'truer' value.

In addition to the analysis of the composite disturbance score between the two groups of *subjects* (which showed a significant difference between them), it was decided to tabulate this score against the *individual words* from the stimulus list. This was designed to answer a question of some importance: Is the tendency towards disturbance on the part of neurotics and controls similar but different in degree? Response to the word-association situation has for many years been held implicitly to be strongly correlated with response to the life situations or 'areas of stress' which the particular words signify. According to the most generally accepted views of neuroticism, one would expect a significant difference in disturbance (both on the test and in the life situation), but a high correlation between the stimuli which promoted the disturbance (both in the test and in the life situation) in the two groups. Many writers, for example, have stressed the view that the neurotic personality is the one who fails to react adequately to the conflicts and difficulties which everyone experiences, but which in most people are in some sense partially resolved. In Table 2 below will be found the disturbance score on words 2–50 for each group.

The rank-difference correlation between the two sets of scores was 0·754, highly significant, showing that there is a strong tendency for the two groups to be disturbed by the same words, though as Table 1 has shown, this disturbance is different in degree in the two populations.

[1] Present writer's italics.

[2] *Loc. cit.* p. 394.

Table 2

Word	Disturbance score: Neurotics	Disturbance score: Controls	Word	Disturbance score: Neurotics	Disturbance score: Controls
Ruler	33	14	Castrate	86	64
Friends	36	18	Wife	47	21
Tired	34	8	Interest	86	40
Vomit	35	24	Murder	37	21
Reason	57	25	Bright	27	9
Intercourse	65	28	Persecution	78	50
Home	32	25	Late	52	18
Parade	36	16	Juice	38	31
Rain	22	7	Children	51	22
Marriage	42	22	Fear	65	45
Fire	34	12	Lamp	26	15
Mother	32	15	Coat	34	15
Nerves	60	35	Kiss	59	33
Naked	37	26	Grace	68	37
Ink	30	15	Germs	41	29
Inferiority	75	39	Father	43	20
Family	39	20	Gravy	41	15
Neutral	76	42	Nice	54	23
Book	28	16	Frame	34	18
Prostitute	52	34	Leap	31	12
Headache	32	21	Colour	49	15
Me	41	13	Venereal	60	33
Circle	28	19	Pain	59	15
Drugs	46	22	Flower	34	13
Dance	45	22			

V. Selective remembering

It will be recalled that at the conclusion of the test, without previous warning the subject was required to note all the stimulus words that he could remember during a 3-min. period.

The number of times each word was recalled was tabulated for neurotic and control groups. The correlation between scores thus obtained was exceedingly high, 0·957, showing that the same selective process was operating in the memory of the members of both groups (i.e. the same words tended to be remembered, and the same words forgotten, by neurotics and controls). As would be expected, the factor of recency played some part, words towards the end of the list being recalled more frequently than words at the beginning. Over and above this factor, however, the greatest tendency present was to remember words connected with family and sex (e.g. intercourse, mother, prostitute, wife, children, kiss, father, venereal, etc.). Approximately 52% of all words recalled pertain to these 'areas of stress', while only about a quarter of the stimulus words can be classified under these two headings.

The next question concerns the relation between the tendency to remember a given word, and motor disturbance on that word. In the case of the neurotics, this correlation was $-0{\cdot}005$, and for the controls $-0{\cdot}082$, thus showing that words which caused motor disturbance tended neither to be remembered, nor to be forgotten or repressed.

The correlation between the number of words remembered by the subjects and the Mill Hill Vocabulary Test scores was 0·549, thus confirming the prediction that the number of words recalled would be correlated with intelligence.

The controls tended to remember more words than did the neurotics (F-ratio: 9·764, significant at the 0·01 level). When the effect of intelligence on the 'number of words

remembered' score was held constant by the part-correlation method, the negative point-biserial correlation with neuroticism decreased to insignificance. Thus the intelligence differences between the two groups account for the greater remembering capacity of the controls.

VI. Summary

This paper describes part of an experiment using the Luria Association-Motor technique. The stimulus list contained fifty words, most of which were designed to have an emotional connotation. A composite score, derived from single measures which were shown to discriminate significantly between groups of neurotic and 'normal' soldiers, was found to have a point-biserial correlation of 0·483 with the criterion, neuroticism. Misclassification on the basis of this composite motor-disturbance score would have been about 25% on the population studied.

Rather high negative correlations of both single and composite motor-disturbance scores with intelligence were obtained; thus, disturbance is not only associated with affective arousals, as previous workers have supposed, but also negatively related with intelligence, in populations which approximate to normal distribution with respect to the latter. That this negative relation with an intelligence test based on knowledge of the meaning of words was not due to the subjects' difficulty in comprehending the meaning of some of the stimuli, was shown by selecting six words on which there was a significant differentiation between the groups (in terms of motor disturbance) and which both on *a priori* and empirical grounds could not have failed to have been understood. The correlation was still, however, negative and highly significant.

It has been shown that Luria's 'functional barrier', which keeps conflict in the higher cortical levels until integrated motor outlets are found, is related not only to emotional stability but also to the general mental ability and efficiency of the organism. This latter factor has not been taken into account by previous workers employing similar techniques.

Analysis of disturbance on the individual stimulus words showed a high correlation ($\rho = 0{\cdot}754$) between the two groups, indicating that the same stimuli tend to produce disturbance in both normals and neurotics, and that the difference between them is one of degree and not of kind.

The same process of selective remembering was apparent in both groups, although the controls remembered a significantly higher proportion of the stimulus words. This difference could, however, be accounted for by the difference in intellectual level between the two groups. There was no relationship between motor disturbance and the forgetting of the word, and thus the concept of repression of disturbing stimuli in this situation was not supported.

References

Albino, R. C. (1948). The stable and labile personality types of Luria in clinically normal individuals. *Brit. J. Psychol.* xxxix, 54–60.

Clarke, A. D. B. (1950). The measurement of emotional instability by means of objective tests. Unpubl. Ph.D. Thesis, Univ. London Lib.

Luria, A. R. (1932). *The Nature of Human Conflicts.* New York, Liveright Inc.

Tendler, A. D. (1945). Significant features of disturbance in free association. *J. Psychol.* xx, 65–89.

Zeller, A. F. (1950). An experimental analogue of repression. I. Historical summary. *Psychol. Bull.* xlvii, 39–51.

(*Manuscript received* 31 *March* 1954)

THE CONSTRUCTION OF AN INDEX OF INTELLIGENCE

By WILLIAM A. BELSON

Audience Research Department, B.B.C.

I. *Introduction* (pp. 44–45); (i) *The originating problem* (pp. 44–45); (ii) *General principle of the device* (p. 45). II. *Method and results* (pp. 45–49): (i) *Stage* 1: *items proposed as correlates of intelligence* (*g*) (pp. 45–46); (ii) *Stage* 2: *item selection* (pp. 46–48); (*a*) *General procedure* (p. 46); (*b*) *Test materials: the criterion test* (p. 47); (*c*) *Test materials: the main items* (p. 47); (*d*) *Results of item selection* (p. 48); (iii) *Stage* 3: *setting up norms* (pp. 48–49). III. *Gathering the necessary information* (pp. 49–50). IV. *Comments on the index* (pp. 50–51): (i) *Some limitations* (pp. 50–51); (ii) *Suggested improvements* (p. 51); (iii) *The principle of indices* (p. 51). V. *Summary* (p. 51). *References* (p. 52).

I. Introduction

(i) *The originating problem*

Intelligence, assessed by a measure primarily of '*g*' (the Vernon Abstraction Test), had emerged in studies by the B.B.C.'s Audience Research Department as a variable highly related to a listener's ability to comprehend broadcast talks (Belson, 1952). Because of this relationship, intelligence (defined as *g*) became an important category for the statement of the results of such studies of comprehension.

In effect, this meant that the most meaningful statement of a programme's comprehensibility was to be made through a joint statement of (*a*) its comprehensibility to people in different intelligence groups, and (*b*) the distribution according to intelligence, of the people who eventually constitute its audience.[1]

The first of these needs could in fact be met in a fairly direct way. Comprehension tests generally involve the playing, in conference rooms, of recorded programmes (of an informative kind) to small, carefully selected groups of people and the subsequent delivery of group tests aimed at finding out if the individuals in the group have been able to grasp the various points presented in the programme. Under these conditions a formal intelligence test could be given with the other tests, so that comprehension scores could in fact be stated in terms of intelligence. It was hard to see, however, how the intelligence distribution of the programme's normal audience could be ascertained. Survey methods would have had to be used in order to meet a representative sample of the programme's actual audience, and this meant that the doorstep interview was the best testing situation that could be expected in a large number of cases—which situation was obviously quite inadequate for the delivery of the usual sort of intelligence test. If survey methods were to be used, what seemed necessary was some short assessor of *g*, simple to use and relatively independent of the conditions of the interview.

It also appeared at the time that it would be a decided advantage if such a device could also be used as a substitute, *in the test room*, for the formal test of *g*. For normal purposes there would have been little or no case for such substitution. The department's studies of comprehension are, however, usually conducted during evening sessions, half of which are devoted to listeners' discussion of their programmes, and the delivery of a formal test

[1] An important additional function of such test results is to allow the matching, in terms of intelligence test score, of groups of people taking part in certain comprehension studies.

had been using up a considerable part of the limited time thus available for experimental work. In addition to this, the listeners who take part are generally unfamiliar with the idea of experiments or of standard testing conditions, and this meant that special preparations for the delivery of the test had normally to be made.

To meet this double need, a short device was developed—a device, that is, to make possible a rapid assessment of *g* under survey conditions, and usable also in the test room. This article presents the device, not so much in its departmental use as for whatever value it may have to social psychology as an independent research tool.

(ii) *General principle of the device*

Two kinds of instrument were considered, one of them a short, disguised, intelligence *test* and the other an *index* of intelligence. The principles on which they work should be carefully distinguished. In the *index*, the basic idea is that if people of different intelligence differ in respect of certain relatively stable characteristics (e.g. education, age, choice of newspaper, occupation), these characteristics may be used as indices of intelligence, the level of intelligence being estimated from them. In the short intelligence *test*, on the other hand, the subject is generally required *at the time of testing* to perform a task in such a way as to demonstrate his ability to see relationships; speed in performance would also be required. While the *short test* is no doubt less open to deliberate falsification than is the *index*, the *test* (as such) is dependent upon the vicissitudes of the interviewing situation: the nature of the training of the interviewer is likely to be important enough, but in addition to this, it is necessary to think seriously of how typical are the intellective performances of people 'caught on the doorstep' as it were; then again there is the difficulty of applying any effective control to the time allowed for the test. The index approach, on the other hand, appears to be relatively independent of the conditions of the interview and does not involve the question of timing. Accordingly, research was directed mainly towards the development of an index.

II. METHOD AND RESULTS

There were three stages in the construction of the index: (1) the compilation of a list of characteristics or items which might with some plausibility be thought to be correlated with intelligence (*g*); (2) the testing of these items against a criterion of intelligence (as defined) in order to select that combination of them yielding the highest multiple-correlation with the criterion; (3) the norming of the index (consisting of the selected items) for the general adult population of Great Britain.

(i) *Stage* 1: *items proposed as correlates of intelligence* (*g*)

The listing of proposed items was restricted to some extent by the requirement that the final index should fit into the general context of one or another of the department's questionnaires and also, to a slight degree, by an emphasis on items likely to be useful in themselves. Those finally listed for study included: occupational level, educational level achieved, age, choice of newspaper, frequency of book reading, frequency of cinema-going, degree of crowding in the home, hobbies and interests, attitude towards the criticism of one's own beliefs (by broadcasters), degree of attachment to radio, attitude towards

listening to the Third Programme, active membership of a society or club, the buying of Penguin or Pelican books, use of the Radio Times in the selection of programmes, general opinion of B.B.C. programmes, likes and dislikes in relation to specific programmes, programme service preference (that is, preference for the Home Service, or the Light Programme, or the Third Programme).

It was of course clear at the outset that occupational level is related to intelligence (Foulds & Raven, 1948; Lingwood, 1952; Vernon, 1949). To make this item apply to the whole adult population, it was proposed to rate housewives by job before marriage and retired people by last job held, though it was realized that this would probably add to the predictive inaccuracy occasioned by the reported wide range of intelligence within any one occupational group. While it also seemed reasonable to expect a useful degree of correlation between educational level and intelligence, the majority of the adult population in Great Britain did not proceed beyond elementary school (leaving at 14 or 15 or earlier) and this suggested that educational level might not in fact be a very discriminating predictor. The inclusion of age in the list had its support in the various studies already made of the relation between age and intelligence test score (Jones & Conrad, 1933; Miles & Miles, 1932; Vincent, 1952), though clearly much would depend upon the nature of the intelligence test (Lorge, 1936; Wechsler, 1950).

(ii) *Stage 2: item selection*

This stage involved: (1) determining the extent of relation to intelligence test score of each of the proposed items; (2) determining, on the bases of these correlations, which combination of the proposed items would give the most accurate prediction of intelligence test score; (3) determining the degree of accuracy to be expected of this prediction.

(*a*) *General procedure.* Some 270 people, constituting a sample representative of the adult population of Great Britain in terms of occupational level, age, and sex, took part in this stage of the work. They were all drawn, by written invitation, from those people who were interviewed previously in one or another of the department's daily surveys of listening and viewing[1] and they met on various evenings in London conference rooms in groups of 15 to 20. While part of the evening was devoted to explaining the requirements of the study and to winning co-operation, the essential steps were: (i) the gathering of information allowing classification of each person in terms of each of the characteristics or items to be examined; (ii) the delivery of a criterion test of intelligence (*g*). All the work was of a paper and pencil kind. Details of the main items and of the criterion test are given below under *Test materials.*

Previous research had established the rates at which people of different age, sex and occupational level would turn up in response to invitation. These differential rates were taken into account in the issuing of invitations but, even so, a group secured by this method of invitation could only approximately fulfil sample requirements; once the group was in the test-room, however, the necessary details (about occupational level, age and sex) were gathered and this made possible the subsequent refinement of the sample so that it finally approximated very closely indeed to the general adult population (in terms of occupational level, age and sex).

[1] These are quota surveys, involving 2500 interviews each day. The controls are social class, age, sex and whether or not gainfully employed.

(*b*) *Test materials: the criterion test.* The test used in this study as a criterion of intelligence was the Vernon Abstraction Test. This test, which consists of forty abstraction-type items (Vernon & Parry, 1949) and imposes a time limit of 15 minutes, has only small verbal and numerical loadings and is primarily a test of *g*. Though it is a paper and pencil test, very little writing indeed is involved in it. It was this test which, as a measure of *g* used in earlier research (Belson, 1952), had brought out the importance of the relation between comprehension and *g*. This fact was one reason for its choice as a criterion in the present study. Another reason, discussed more fully later on, was that its use here made it possible to draw directly from a large volume of data gathered in earlier research.

(*c*) *Test materials: the main items.* Of the twenty-one items proposed as components of the index, only occupational level, age, educational level and 'frequency of book reading' are described in any detail here, as these were in fact the four items yielding the highest multiple-correlation with the criterion (details of the other items, along with their correlations with the criterion are available from D.S.I.R. Technical Information and Documents Unit).

The scaling of occupational level required the collection of details of the subjects' occupational background. Thus each of the subjects who was in a job completed a brief occupational questionnaire giving the name of his or her job, a short statement (of not more than two lines) of what was involved in the job, qualifications (if any) required for the job, and the number of people (if any) working under his or her authority. Housewives gave the same information for job before marriage and retired people referred to last full-time job before retirement. This information made it possible later on to place each person in one or another of seven occupational groups scaled according to 'skill and training required for the job'. These groups were: higher professional and administrative; lower professional, technical and executive; highly skilled; skilled; moderately skilled; semi-skilled; unskilled. On this scale each group is described in terms of skill and training required and is illustrated by sample occupations. The construction of the scale, which had its beginnings in one presented by Spielman & Burt (1926), was influenced by research data giving mean intelligence test scores in various occupations (Foulds & Raven, 1948; Lingwood, 1952; Vernon, 1949).

For the scaling of educational level, the following educational information was recorded: age of ceasing full-time education, type of school attended, the nature of any (further) education completed since leaving school, any examinations passed or certificates gained. This information then made it possible to place each person in one or another of eight educational groups scaled according to 'general (or academic) educational ability or equipment'. (Both the occupational and the educational scales are available from the D.S.I.R. Technical Information and Documents Unit.)

Age was recorded directly on a personal detail form. Classification according to 'frequency of book reading' was based on the answer to the question 'have you managed to read any books recently?', the subject answering 'yes', 'no' or perhaps giving no reply at all; this item is a very close approximation to one used successfully by Goossen in his American study (Goossen, 1950). With the exception of 'choice of newspaper' (where the names of newspapers read were given), interests and hobbies (where a list of these was volunteered) and one or two other items, the classifications were based on multiple-choice responses to questions of this kind.

(*d*) *Results of item selection.* In the final analysis the more highly predictive of the individual items were occupational level (+0·65), educational level (+0·57), age (−0·48), book reading (+0·38) and 'attachment to radio' (+0·35).[1] These correlations and the inter-item correlations are shown in Table 1. Occupational level and age yielded a multiple-correlation with intelligence of +0·785, educational level increasing it to 0·798 and 'book reading' increasing it to 0·808. It should be noted that while educational level is the second most powerful *individual* predictor, it is quite highly correlated with occupational level; age, on the other hand, has practically zero (−0·07) correlation with occupational level. Accordingly it is not surprising that, for the purpose of maximizing correlation with the criterion, age is a better item to use in conjunction with occupational level than is educational level.

Table 1. *Criterion and inter-item correlations*

	Intelligence (*g*)	Occupation	Education	Age	Book reading
Occupation	+0·65	—	—	—	—
Education	+0·57	+0·64	—	—	—
Age	−0·48	−0·07	−0·15	—	—
Book reading	+0·38	+0·28	+0·23	−0·14	—
Attachment to radio	+0·35	+0·38	+0·37	−0·27	+0·53

From a practical point of view, there seemed much to be gained by limiting the index to occupational level and age and sacrificing the small gain in validity coming with the use of the third and the fourth items for the simplicity of a two-item index. Accordingly, only occupational level and age were retained.

(iii) *Stage* 3: *setting up norms*

Norming was carried out in two steps. In the first step, the test results of an approximately rectangular sample of people (that is, a group with an approximately equal number of people in each relevant subsection) was used to establish the mean intelligence test scores in each of the various occupation/age groups (e.g. skilled, 26–35 years; unskilled, 66+years). The bulk of the material for doing this was fortunately available from an earlier study in which each of over 1000 people of widely different age, occupational level and intelligence, completed both the intelligence test and the occupation/age detail form used in the present study (Belson, 1952). There were six age groups at each of six occupational levels, making thirty-six occupation/age subgroups in all. The number of people in each of these averaged about 30. Mean intelligence test scores of the six subgroups at each occupational level were plotted and a limited amount of smoothing was done. The thirty-six adjusted mean intelligence test scores are set out in Table 2.

Secondly, appropriate questions about occupational level and age were included in a random survey of the adult civilian population of Great Britain (approximately 4700 interviews) being carried out at the time, thus making it possible to determine the percentage of the population falling into each of the thirty-six occupation/age groups. These percentages appear in Table 2.

[1] Those people answering 'yes' to the question about book reading tended to be more intelligent than those answering 'no' or giving no reply; those using extreme statements on the attachment scale or giving no reply tended to be less intelligent than those choosing the more moderate statements.

Since the percentage size and the mean test score of each group was known, it was then possible, through a simple cumulative frequency procedure, to determine which occupation/age groups fell into the most intelligent 10% of the population, which fell into the next 10% group, which into the third 10% group, and so on right down to the lowest scoring 10% group.

Table 2. *Percentage size of each occupation/age group, mean test scores and decile groups*

Age group	Occupational level* 1/2	3	4	5	6	7	Others	Percentage adult population of Great Britain† (excluding servicemen)
21–25 years								
Mean test score	34·9	33·8	29·2	23·5	21·0	18·0	—	—
% G.B. population	0·1	0·8	1·5	2·0	1·2	0·7	0·2	6·5
Decile group	I	I	II	III	IV	VI	—	—
26–35 years								
Mean test score	33·8	31·8	27·4	22·4	19·5	16·0	—	—
% G.B. population	0·5	1·9	5·7	4·5	4·3	1·9	0·6	19·4
Decile group	I	I	II	IV	V	VII	—	—
36–45 years								
Mean test score	33·4	31·3	24·8	19·8	17·3	16·0	—	—
% G.B. population	0·9	2·7	5·0	6·5	4·8	3·0	1·0	23·9
Decile group	I	I	III	V	VI	VII	—	—
46–55 years								
Mean test score	30·7	29·6	22·2	17·3	15·5	13·8	—	—
% G.B. population	0·8	2·0	5·3	5·2	4·5	3·2	1·1	22·1
Decile group	I	I	IV	VII	VIII	IX	—	—
56–65 years								
Mean test score	28·0	27·4	18·8	13·8	12·5	10·2	—	—
% G.B. population	0·8	1·1	2·4	3·3	3·1	2·8	1·1	14·6
Decile group	II	II	VI	VIII	IX	IX	—	—
66 + years								
Mean test score	22·7	21·5	13·0	10·0	9·0	7·5	—	—
% G.B. population	0·6	0·8	2·3	3·6	2·4	2·2	1·6	13·5
Decile group	III	IV	IX	X	X	X	—	—
All ages (totals)								
% G.B. population	3·7	9·3	22·2	25·1	20·3	13·8	5·6	100·0

* 1/2 = higher and lower professional; 3 = highly skilled; 4 = skilled; 5 = moderately skilled; 6 = semi-skilled; 7 = unskilled; others = persons without occupational record or unclassifiable.

† Based on the random sample referred to on p. 48.

The ten groups thus established were numbered, I, II, III, ..., X and were called the first, second, third, etc., decile groups. The decile groups into which the various occupation/age combinations fell are entered in Table 2 in Roman type and this set of entries may be used without further reference to any of the other entries to estimate intelligence level from occupational level and age. (In normal practice Table 2 would be reduced to a simple two-way table with only decile groups entered in the cells.) From Table 2 it will be seen that if a person interviewed in a survey is of occupational group 5 and of age group 46–55 he will be rated as falling into the seventh decile group (VII) in respect of intelligence.

III. Gathering the necessary information

In gathering (during a survey interview) the information required for use of the index, questions about age and occupation may be put towards the end of the interview. Age should be asked for directly and questions about occupation should be as follows.

(*a*) What is your job? .....
If *housewife* or *retired*, ask:
What was your job before you married*/retired? .....
.....
If informant names more than one job, record the last one.

(*b*) What kind of work do/did you do in this job? .....
.....
In some cases (e.g. bus conductor, railway porter) question (*b*) will not be necessary.

(*c*) Do/did you happen to have anyone working under your authority in your job?.....
If 'yes', ask:
How many?.....

(*d*) What training or qualification did you need for the job? .....

* If housewife and no previous occupation, refer all these questions to husband's job.

The interviewer merely records the information without making either rating or judgement. Ratings are made by office staff trained in the use of the scale, and of course, the more extensive their knowledge about jobs, the better their ratings are likely to be. In test room sessions the same information can be collected quite easily through a standard form.

IV. Comments on the index

(i) *Some limitations*

It is important to point out, first of all, that this index was designed as a classifying device for the *grouping of results*, and that though it works *through* prediction of individual intelligence, it is not meant for the prediction of *individual* intelligence as such. Though it might be of some use as a general screening device, it is essentially a tool for sociological inquiries involving the grouping of results.

Secondly, it should be pointed out that the multiple-correlation of +0·785 is based upon information gathered under test room conditions and not in doorstep interviews, even though the index is required to be valid under such doorstep conditions. Had the items selected been more of a 'test' nature this might have been crucial, but with variables like occupation and age it was tentatively assumed that there would not be appreciable differences in the information collected under the two conditions. Further research on the point should not, however, be ruled out. Somewhat more serious, perhaps, is the limitation arising out of basing the index on the responses of a London sample, for there is evidence (Vernon, 1947) to suggest that some regional differences, small but significant, may be expected in mean intelligence test scores within a given occupation. If this is so, it means that in national surveys some regional adjustment is necessary.

Thirdly, there is no point in using the present index, viewed as a general research tool, unless g is in fact related in an important or relevant way to the variable under study. Its use in studies of comprehension (of broadcast talks), for instance, arose out of the high degree of correlation found to exist between g and comprehension score, a relation probably accentuated by the essentially non-stop nature of the radio talk. It is worth noting in this connexion, however, that g is not necessarily the most powerful single predictor of comprehension available. Conceivably an intelligence test with an appreciable verbal loading and delivered at speed would correlate with comprehension score at a slightly higher level than does a test primarily of g. Unfortunately, a test of this other

kind, really suitable for the very special test-room conditions concerned, was not available at the time of the collection of the bulk of the data used in this project—namely the set of records giving age, occupational level and a measure of *g* referred to above (Belson, 1952).

(ii) *Suggested improvements*

No doubt a number of improvements could be made in the index. In the first place it should be possible to continue with the adjustment of the scale in accordance with research findings which give mean intelligence test scores in specific occupations. How far this could be carried without destroying the basic dimension of 'skill and training required' is of course a matter of some importance, and in the end depends upon the relation between that variable and intelligence. The trend in such an adjustment is clearly towards a replacement of the 'skill and training' scale by a direct intelligence scale, each section being copiously illustrated by occupations for which the mean intelligence test score has been determined by research.

Another improvement might be made, perhaps, through using educational information in the interpretation of the occupational details provided; again, one or two questions about occupation might be added (e.g. whether monthly or weekly paid) to increase discrimination between jobs otherwise similar in description; and of course the better the fund of general occupational information, the more accurate will be the placement of specific occupations upon the scale. Again, additional items (i.e. items additional to the two actually used) might be added to the index, though there is probably only a narrow margin within which this process could in practice be continued.

(iii) *The principle of indices*

Quite apart from the value which this index of *g* may have as an independent research tool, it does seem possible to employ the principles used in its construction to achieve rapid assessment under survey conditions (and at a level of accuracy suitable for the grouping of results) of a certain range of other abilities and variables, the measurement of which normally requires the administration of relatively lengthy tests under carefully controlled conditions.

V. SUMMARY

For certain purposes relating to audience research it was found necessary to develop a simple method of estimating intelligence which would be independent of the administration of conventional tests. To this end an 'index' consisting of grades on a ten-point (decile) scale was evolved. The index in its final form was based upon two items, 'age' and 'occupational level', which together yielded a multiple correlation of +0·785 with intelligence (defined as *g* and assessed by the Vernon Abstraction Test). Occupational level, which is here used in a specialized sense, refers either to present job or job before retirement or, for housewives, to job before marriage. A simple two-way table was developed for the direct conversion of age and occupational level into an intelligence grade. The index as constructed and normed is intended for use in Great Britain.

References

Belson, W. A. (1952). The comprehensibility of 'Topic for Tonight'. An Audience Research Department (B.B.C.) report, LR/52/1080.

Foulds, G. A. & Raven, J. C. (1948). Intellectual ability and occupational grade. *Occup. Psychol.* xxii, 197.

Goossen, C. V. (1950). The Goossen hidden intelligence test. *Pub. Op. Q.* xiv, 759.

Jones, H. E. & Conrad, H. S. (1933). The growth and decline of intelligence: a study of a homogeneous group between the ages of ten and sixty. *Genet. Psychol. Monog.* xiii, 223.

Lingwood, J. (1952). Test performances of A.T.S. recruits from certain civilian occupations. *Occup. Psychol.* xxvi, 35.

Lorge, I. (1936). The influence of the test upon the nature of mental decline as a function of age. *J. Educ. Psychol.* xxvii, 101.

Miles, C. C. & Miles, W. R. (1932). The correlation of intelligence scores and chronological age from early to late maturity. *Amer. J. Psychol.* xliv, 44.

Spielman, W. & Burt. C. (1926). The estimation of intelligence in vocational guidance. In *A Study in Vocational Guidance.* Industrial Fatigue Research Board Rep. No. 33.

Vernon, P. E. (1947). The variations of intelligence with occupation, age and locality. *Brit. J. Psychol. Statist. Sect.* i, 52.

Vernon, P. E. (1949). Occupational norms for the 20-minute Progressive Matrices Test. *Occup. Psychol.* xxiii, 58.

Vernon, P. E. & Parry, J. B. (1949). *Personnel Selection in the British Forces.* London: University of London Press.

Vincent, D. F. (1952). The linear relationship between age and score of adults in intelligence tests. *Occup. Psychol.* xxvi, 243.

Wechsler, D. (1950). *The Measurement of Adult Intelligence.* Baltimore: Williams and Wilkins Co.

(*Manuscript received* 29 *April* 1954)

SPEECH-BREATHING ACTIVITY—A MEASURE OF TENSION AND AFFECT DURING INTERVIEWS

By FRIEDA GOLDMAN-EISLER

Medical Research Council Unit for Research in Occupational Adaptation, Maudsley Hospital, London

I. *Introduction* (pp. 53–54). II. *The problem* (p. 54). III. *The measures* (pp. 54–56): (i) *Method of recording breath rate during speech* (p. 54); (ii) *Measuring speech production* (pp. 54–56). IV. *Consistency and variability of measures* (pp. 57–60): (i) *Material* (p. 57); (ii) *Results* (pp. 57–60). V. *Discussion* (pp. 60–62): (i) *Psychological significance of breath rate* (pp. 60–61); (ii) *Psychological significance of the number of syllables per expiration* (pp. 61–62). VI. *Summary* (p. 62). *References* (p. 63).

I. Introduction

Interviewers of all kinds, especially psychotherapists and psychiatrists, have for long been beset by the problem of how to assess the fluctuations of tension and affect which take place in persons while being interviewed. It is increasingly being realized (Mowrer, 1953; Ruesch & Prestwood, 1949) that the best way to study such changes is through indicators of mental and emotional states measured during, and by way of, the speech process itself. The present study is meant to contribute to the solution of this problem.

In a search for such indicators the writer (1954*a*) studied time sequences, rhythm, and rate of speech, as well as a number of grammatical categories in language obtained through time and speech recordings. These measures were tested for consistency within, and sensitivity in discriminating between individuals, as well as between situations (Goldman-Eisler, 1952, 1954*a*). The speech rate, or speed of talking, was found to be the most consistent and most sensitive of these measures, and variations in the rate of speech were found to be indicative of a total change in linguistic activity. The changes, which included content and affected both speakers, were of a compensatory and self-regulating nature. This finding suggested that variations in the speed of talking may serve to restore the balance of the linguistic equilibrium where this has been disturbed as a result either of strong affect or of tension having been aroused during the interview. As a consequence of these studies the rate of speech production was subjected to further detailed analysis (Goldman-Eisler, 1954*b*).

So far the following facts concerning speech rate have emerged. Speech rate is relatively constant for any person from one conversation to another when calculated as an average over the total length of an individual's speech output. When speech rates were calculated for the individual utterances in a conversation (as the average number of syllables uttered each minute) consistency was confined to long utterances containing 100 syllables or more; for shorter utterances a person's speed varied over five times as wide a range as for long utterances. Moreover, short utterances were, on the average, faster than long utterances. The fast rates of speech which were common in short utterances rarely occurred in utterances more than 60 syllables long, and never in utterances of more than 100 syllables. The greater speed of short utterances was assumed to be due to freedom from inhibiting factors, both physiological (need to take breath) and psychological

(emotional blocking or need to think), which are bound to slow down speech to some extent. The psychological implications of speed and acceleration in speech have been discussed elsewhere (Goldman-Eisler, 1954*b*).

II. The problem

The present paper deals with the attempt to break down the measure of speech rate into constituent measures derived from respiratory behaviour during speech. In view of the importance of breathing to the production of speech on the one hand, and of the emotional significance of respiratory variations on the other, measures of speech-breathing activity promised to be more immediately relevant to the changing states of tension and affect during interviews than speech rate unqualified by further definition could be expected to be.

The problem to be solved was twofold. A workable method of measuring breathing during speech had to be found, and such measures of speech-breathing activity had to be developed as would prove psychologically relevant. The writer's object in the present paper is primarily to call attention to the technique.

III. The measures

(i) *Method of recording breath rate during speech*

The recording of breathing by the usual apparatus, pneumograph, spirometer, etc., seems most undesirable if the speakers are to remain as uninhibited and spontaneous as is necessary if we want to make investigations about natural and spontaneous speech, and use our type of analysis in psychiatric interviews and psychotherapy.

The writer discovered that by listening carefully she could hear quite distinctly the noise produced by speakers when inhaling air. Two observers counting inhalations of recordings, for which verbatim manuscripts were at hand, were practically in complete agreement. The reliability of this method, of course, presupposes a highly sensitive microphone and most conscientious and accurate listening.

(ii) *Measuring speech production*

Using the above method the following measures for each utterance are at our disposal:

(1) Duration of utterance, t.
(2) Number of syllables in each utterance, N_s.
(3) Number of inspirations in each utterance, N_i.

From these we derive:

(4) The rate of respiration, i.e. the number of inspirations per minute, $N_i/t = RR$ (or *rr* for low rate of respiration).
(5) The rate of speech, i.e. the number of syllables per minute calculated for each utterance, $N_s/t = SR$.
(6) Output per breath, i.e. the number of syllables per expiration (or rate of expulsion of syllables), $N_s/N_i = ER$ (or *er* for low expulsion rate).
(7) Its reciprocal, measuring the proportion of the returned air current per syllable spoken expressed as a percentage, $N_i/N_s \times 100$ ($V\% =$ ventilation percentage).

From the above it may be seen that the rate of speech N_s/t is the product of breath rate and output per breath, i.e. $N_i/t \times N_s/N_i = N_s/t$. From this formula it follows that there could be an infinite variety of combinations between the two factors N_i/t and N_s/N_i possible for any particular rate of speech, the number of syllables per expiration decreasing as breath rate increases and vice versa. Each particular rate of speech is a hyperbola, the two sides being breath rate and number of syllables per breath. Thus when a person speaks at a rate, e.g., of 300 syllables per minute, he may have achieved it by inhaling 10 times and uttering an average of 30 syllables between each inspiration, or 30 times speaking only 10 syllables between each breath, or anything between as well as beyond these figures.

Thus the same rate of speech would cover very different physiological and psychological processes operating in the production of it. This is reminiscent of the different moods which may be covered by the same metronomic tempo in music.

We are proposing to describe any utterance in respect of the two processes which combined give us a measure of the speech rate, namely the respiratory activity and the degree of its utilization as speech sounds. The former is given in the rate of respiration (number of inspirations per minute) and the latter as the number of syllables per expiration being, as will be shown later, also a measure of the rate of escape or expulsion of the returning (expired) air current. Fig. 1 with the speed hyperbolas (on double-logarithmic paper they are straight lines) drawn in allows us to locate each utterance provided we know the number of syllables, the number of inspirations contained in it, as well as its duration. Examples of graphical representation on double-logarithmic paper are on p. 56, and we can see that even within the narrow confines of constant speed, utterances differ widely in respect of breathing activity (measured by *RR*) and the economy of distributing it per syllable. Speed of talking and rate of respiration are independent dimensions of speech, at least in general, although on occasions we meet with speech situations in which acceleration was achieved by one kind of mechanism chiefly, either by increasing or decreasing the rate of respiration. Both cases are possible, and we may expect each to have different psychological connotations.

Thus utterances can be plotted as is illustrated in Figs. 1 and 2.

The scheme in Fig. 1 represents the various ways of accelerating utterances. In the left bottom corner of the rectangle ($rr \times er$) we would place utterances produced with minimal supply of air and minimal output per expiration; they would thus be utterances of minimal speed. Moving vertically upwards we would find utterances with increasing output of syllables per expiration carrying minimal ventilation, while the bottom horizontal side of the quadrangle would be for utterances of increasing air supply and minimal output of syllables per expiration. As utterances move from there along the right vertical towards the right top corner ($RR \times ER$) we have reached the maximum speed achieved by maximum breathing activity, and maximum air expenditure. If for this scheme we use double-logarithmic paper and plot each utterance according to respiration rate (abscissa) and output of syllables per expiration (ordinate) the rate of speech will be the diagonal while the two ordinates RR and $N_s/N_i = ER$ are two sides, according to the above formula.

Such graphical presentation of utterances in any one conversation illustrates at first sight the basic differences in the nature of speech production even where there are no differences of speech rate (Fig. 2).

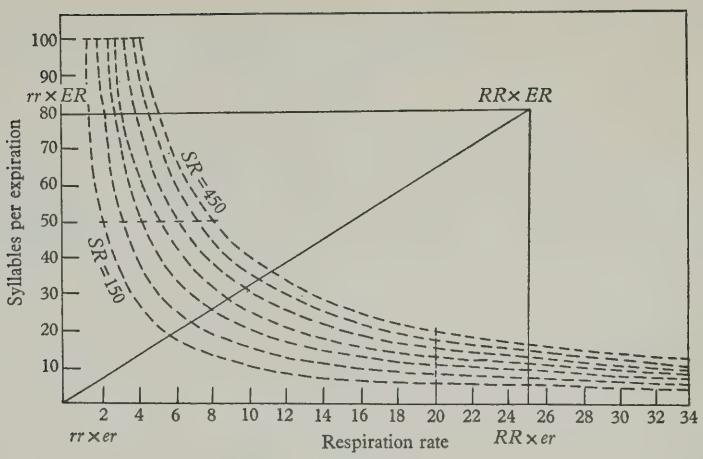

Fig. 1. The broken lines indicate the speech rates (*SR*) rising by 50 from 150 to 450 syllables per minute achieved at respiration rates up to 34 a minute and an output of up to 100 syllables per expiration.

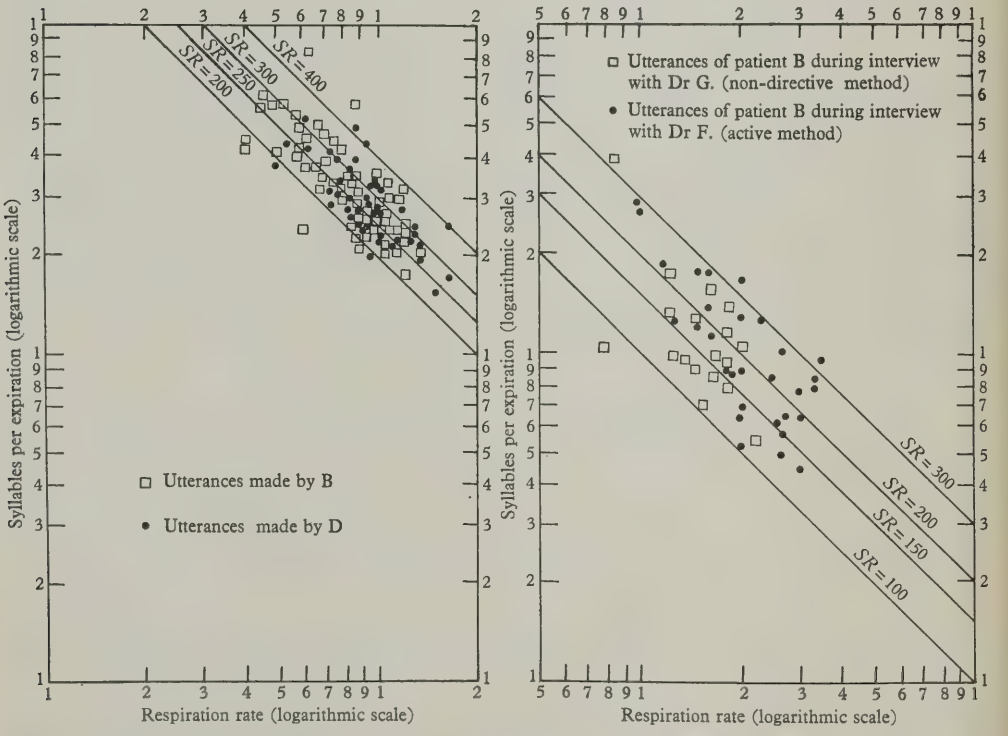
Syllables per expiration (logarithmic scale)
SR = 300
SR = 250
SR = 200
SR = 400
Utterances made by B
Utterances made by D
Respiration rate (logarithmic scale)
Utterances of patient B during interview with Dr G. (non-directive method)
Utterances of patient B during interview with Dr F. (active method)
SR = 300
SR = 200
SR = 150
SR = 100

Fig. 2. On the left are plotted the utterances of two persons in conversation whose speech rates (*SR*) were about equal, but who differed in output per expiration and respiration rate. The logarithmic scales indicate respiration rates of 1–20 on the left, and 5–100 on the right, and for syllables per expiration of 1–100 on both sides. On the right are plotted utterances from two interviews of a schizophrenic patient B with two different psychiatrists. Dr G. used the non-directive, and Dr F. the active technique of interviewing. The two interviews are differentiated not so much by speech rates as by the higher degree of ventilation with which B responded in the active interview.

IV. CONSISTENCY AND VARIABILITY OF MEASURES

The above observations and deductions provide information as to the possible relations in speech between breath rate, number of syllables per expiration, and speech rate. The individual examples illustrated how these possible combinations may appear in actual situations in some individuals.

The 'goodness' or usefulness of the measures introduced depends on their capacity for measuring individual differences.

Tests of consistency and discriminative capacity of the rate of respiration in speech, as well as of the number of syllables per expiration (or its reciprocal), carried out by means of the statistical method of analysis of variance, led to the following results.

(i) *Material*

The material consisted of twelve discussions between four adults of superior intelligence in different combinations, seven interviews between the experimenter and four young female neurotics, one schizophrenic, and three research workers on the staff, as well as four diagnostic interviews by a psychiatrist with patients attending the outpatient clinic. Altogether, twenty-three interviews containing 829 utterances served as the material for this investigation.

The number of syllables, as well as of inspirations, were counted and speech and breath rate calculated, as well as the output of speech (number of syllables) per breath and the reciprocal of this value, the ventilation percentage calculated.

(ii) *Results*

Two Analyses of Variance were carried out. The one was based on the four individuals who had discussed controversial subjects with each other. Three recorded discussions were available from each of them, and it was therefore possible to test their individual consistency for different conversations with different interlocutors, on different occasions, and discussing different topics, and to compare the variance within each of these individuals with the variance of the tested measurements from person to person. This group was a comparatively homogeneous one consisting of research workers and academic personnel. The second Analysis of Variance was based on utterances derived from a mixed group of twelve individuals, patients and normals, from whom only one interview-recording was obtained. The internal consistency for these measurements was obtained by comparing those taken in the first half with those from the second half of the interview.

Tables 1–4 show the means and standard deviations for these two groups of respiration rate and of the degree of ventilation as measured by the reciprocal of the number of syllables per expiration. (The reciprocal was chosen for statistical calculation, being more normally distributed.)

These data were obtained by two interviewers; the table of means shows that the differences cannot be due to the interviewer's influence and that the subjects interviewed by the experimenter (*Ma*, *Nu*, *Stn*, *S.O.*, *S.* I, *S.* II) differ much more among themselves (this group includes patients as well as normals) than the subjects interviewed by *B* who are all patients. Neither are systematic differences between the patients and the normals evident, but the groups are too small to answer the question of such differences.

It is clear from these tables that not only the rate of respiration but also the degree of ventilation per speech unit, or reciprocally the number of syllables spoken per expiration,

are measures of high consistency and capacity of discriminating between individuals. At the same time they are sensitive to fluctuations in speech-breathing behaviour, as the significant differences between the first and second halves of the interviews in the second set of measurements (Tables 3 and 4) show.

Table 1. *Means and standard deviations of respiration rates of four intelligent normal adults based on* 463 *utterances made in twelve conversations* (*three per person*)

	Persons											
	A			*B*			*C*			*D*		
Conversations	Mean	S.D.	*N*	Mean	S.D.	*N*	Mean	S.D.	*N*	Mean	S.D.	*N*
1	6·06	2·59	36	8·42	2·63	43	8·81	5·02	59	8·48	2·43	59
2	5·66	2·31	13	6·25	2·50	17	9·16	4·66	42	9·76	2·65	48
3	7·63	3·97	29	8·32	3·62	24	9·95	3·56	31	8·27	3·20	62

Analysis of variance

Source	Degrees of freedom	Sum of squares	Mean square
Persons	3	430·89	143·63
Conversations	8	10·24	1·28
Error	451	5602·86	12·13
Total	462		

F, persons/error = 11·84. *P* at 0·001 = 5·42 (D.F. 3∞). $\epsilon^* = 0{\cdot}957$.
F, persons/conversations = 112·21. *P* at 0·001 = 15·83 (D.F. 3 and 8).
* Epsilon is a measure of correlation based on the analysis of variance (see Peters and Van Voorhis, *Statistical Procedures and their Mathematical Bases*, 1940, pp. 325, 353. New York: McGraw-Hill).

Table 2. *Means and standard deviations of ventilation percentages of four intelligent normal adults based on* 463 *utterances made in twelve conversations* (*three per person*)

	Persons											
	A			*B*			*C*			*D*		
Conversations	Mean	S.D.	*N*	Mean	S.D.	*N*	Mean	S.D.	*N*	Mean	S.D.	*N*
1	2·90	1·13	36	3·48	1·18	43	5·14	3·23	59	3·19	0·98	59
2	2·28	1·06	13	3·54	1·84	17	5·04	2·47	42	3·59	1·00	48
3	3·08	1·39	29	3·52	1·18	24	5·59	3·23	31	3·40	1·42	62

Analysis of variance

Source	Degrees of freedom	Sum of squares	Mean square
Persons	3	388·66	122·89
Conversations	8	16·17	2·02
Error	451	1694·62	3·67
Total	462		

F, persons/error = 33·49. *P* at 0·001 = 5·42 (for D.F. 3∞). $\epsilon = 0{\cdot}985$.
F, persons/conversations = 60·84. *P* at 0·001 = 15·83 (D.F. 388).

Comparing the capacity of respiration rate and ventilation percentage to discriminate between individuals both analyses of variance show the latter measure to be more sensitive in discriminating between persons (the *F* ratios for *V*% being 126·3 and for *RR* 29·6), and the second analysis of variance (Tables 3 and 4) also shows that the ventilation percentage discriminates more sharply ($F = 43{\cdot}3$, *P* beyond 0·001) between the first and second part of the interview than the respiration rate ($F = 2{\cdot}62$); the reliability coefficients ϵ calculated on the basis of variances were for the rate of respiration 0·957 and

Table 3. *Means and standard deviations of respiration rates of nine patients and three intelligent normal adults based on 355 utterances made in thirteen interviews*

Persons		First half of interview			Second half of interview		
Patients		Mean	S.D.	*N*	Mean	S.D.	*N*
P III (male)	interviewed by *B*	5·15	2·28	8	4·05	1·81	8
P VII (male)		9·12	3·59	11	10·48	3·32	11
P IX (male)		7·52	2·06	10	6·83	2·04	10
P X (male)		8·61	1·62	12	7·38	2·85	12
Ma (female)	interviewed by *A*	10·46	2·56	17	9·12	2·53	17
Nu (female)		9·78	2·18	24	9·46	1·72	24
Stn (female)		8·83	3·48	18	7·09	4·04	18
G (schizophrenic)		10·57	3·19	34	9·32	2·30	34
Normals							
S.O. (male)		3·88	1·37	6	7·10	3·62	5
S. I (female)		7·76	1·79	18	6·61	1·43	18
S. II (male)		9·49	3·81	20	8·33	2·39	20

Analysis of variance

Source	Degrees of freedom	Sum of squares	Mean square
Between persons	10	839·59	83·96
Between first and second half of interview	11	82·09	7·46
Error	334	947·77	2·84
Total	355		

F, persons/error = 29·56. *P* at 0·001 = 2·74. $\epsilon = 0{\cdot}983$.
F, between first and second half of interview/error = 2·62. *P* at 0·001 = 2·74.
F, persons/conversations = 11·25. *P* at 0·001 = 7·63.

Table 4. *Means and standard deviations of ventilation percentages of nine patients and three intelligent normal adults based on 356 utterances made in thirteen interviews*

Persons		First half of interview			Second half of interview			
Patients		Mean	S.D.	*N*	Mean	S.D.	*N*	dM_1–M_2
P III (male)	interviewed by *B*	2·76	1·11	8	2·04	0·95	8	0·72
P VII (male)		3·56	1·07	11	4·12	1·40	11	2·16
P IX (male)		3·46	1·14	10	3·02	1·12	10	0·44
P X (male)		3·32	0·71	12	2·79	0·96	12	0·53
Ma (female)	interviewed by *A*	4·75	1·43	17	3·75	1·18	17	1·00
Nu (female)		4·93	1·13	24	4·81	1·30	24	0·12
Stn (female)		3·64	1·30	18	4·15	3·32	18	0·51
G (schizophrenic)		4·67	1·51	34	3·58	0·87	34	1·09
Normals								
S.O. (male)		2·08	0·70	6	2·80	1·16	6	0·72
S. I (female)		3·15	1·04	18	2·84	0·68	18	0·31
S. II (male)		5·13	1·93	20	4·35	0·72	20	0·78

Analysis of variance

Source	Degrees of freedom	Sum of squares	Mean square
Between persons	10	921·88	92·19
Between first and second half of interview	11	347·92	31·63
Error	335	244·44	0·73
Total	356		

F, persons/error = 126·3. *P* at 0·001 = 2·74. $\epsilon = 0{\cdot}996$.
F, between first and second half of interview/error = 43·33.
F, persons/conversations = 2·72 (*P* 0·05).

0·983 and for the ventilation percentage 0·987 and 0·996. This was in both cases due to the smaller error of the ventilation percentage, i.e. to the greater constancy of this measure; the comparison of the first and second part of the interviews shows that the trend of the change is clearly in the direction of a decreasing degree of ventilation as the interview moves on. This may be explained when we assume that normally interviewees settle down after a state of initial excitement or flurry to a more calm and better controlled state of mind. It may be different with resistant and rigid interviewees loosening up towards the end, as was the case with *S.O.*, or when a more stressful topic or topics of greater emotional import crop up only in the second part of the interview, as was the case with *Stn* when the subject of pregnancy and child-bearing touching upon the patient's anxieties was discussed.

Correlations between respiration and speech rate, as well as between speech rate and output of speech per expiration, were not calculated. The reason was that observation showed that within limits all kinds of combinations occurred—and this can be seen from Fig. 2—as was envisaged on the basis of the mathematically possible mutual and reciprocal variations of the two variables in speech rate. The latter tended sometimes to be accelerated more by increasing the output per expiration, and the respiration rate varied little; this state of affairs would yield a positive correlation of speech rate and output of speech per expiration, and an insignificant correlation of speech rate and respiration rate. In some other cases the more responsible factor for the acceleration of speech would be an increase in the rate of breathing, with speech output per expiration varying within more narrow limits.

These proportions of the two factors in speech rate varied not only in the same person, but from utterance to utterance within the same conversation. Clear tendencies in one or the other direction were not usual in our sample, although where they appear they would classify a conversation as a whole according to the processes involved in its production.

V. Discussion

The varying proportions in which respiration rate and output of syllables per expiration combine in speech rate indicate that different states of the organism may underlie the same rate of speech production. We may therefore expect also different psychological states to go with the different ingredients of breath rate, and of speech output per breath (number of syllables per expiration).

(i) *Psychological significance of breath rate*

Experimental evidence, as reviewed by Clausen (1951) and Altschule (1953), supports the conclusion that while respiratory effects vary markedly in degree and character, hyperventilation is the common accompaniment of strong emotion and mental disturbance. Studies of the relation of respiration and thinking (Golla, 1921; Schilling, 1929; Suter, 1912) show a picture of reduced breathing activity, breath holding during moments of attention, delays, or irregular breathing. Psychoanalytic writers have frequently stressed the relation between anxiety and breathing (Fenichel, 1953; Freud, 1936; Harnik, 1930; Kubie, 1948). Thus Fenichel (1953) reports that: 'Analysts daily observe passing stoppages of breathing going along with an increase in castration anxiety.... The impression is given that, in inhibiting breathing at an intended new action, at a

perception, at a change of attention, the ego tests out whether or not it needs to be afraid.'

The amount of respiratory activity during speech should therefore be expected to indicate strength of affect as far as it finds expression in speech. Easy and free-flowing affect would manifest itself in ample ventilation, excitement in hyperventilation, whereas repressed affect, attention, caution, and fear would be revealed in inhibited breathing.

The rate of respiration by telling us how frequently the tidal air has to be restored informs us about the amount of breathing activity. Although to make this information complete the further datum of the volume of tidal air is required, the rate of respiration has been found to be a relatively efficient measure of ventilation (Altschule, 1953; Ax, 1953), and more significant in discriminating between emotional states than an index of volume composed of rate times amplitude (Ax, 1953). Amplitude alone or the Inspiration/Expiration ratio were not significant at all.

(ii) *Psychological significance of the number of syllables per expiration*

The following facts from the physiology of speech production concern us here: (*a*) the energy necessary for the production of articulate sound is supplied by the expiratory phase of the breathing cycle; (*b*) the current of expired air can be modified at will, owing to the cerebral connexion with the respiratory centre; and (*c*) ordinarily expiration is a mere cessation of the active process of inspiration. While the latter cannot be much prolonged—it being normally checked by the Hering-Breuer reflex—expiration takes place spontaneously and requires only to be regulated to form a continuous current of air of sufficient duration for the production of sound.

It is possible to increase this outgoing current of air either by expelling a stronger current of air following a moderate intake of breath, by means of special expiratory activity involving muscular pressure, and so emptying the lungs, or by taking a deeper breath and so filling the lungs with a greater volume of air. An increased activity of the respiratory musculature of a voluntary and controlled kind is therefore necessary for both emphatic speech and acceleration.

Two aspects of the respiratory activity are thus relevant to speech production, namely (Meyer, 1892): (*a*) the rate of escape or expulsion of the outflowing current of air, or reciprocally the degree of ventilation of speech (i.e. the amount of air escaping through speech); and (*b*) the neuro-muscular activity controlling the action of the respiratory musculature and thus regulating the rate of escape of the expelled air current.

We are able to modify the strength and duration of the outflowing current of air, allowing it either to escape slowly and evenly as in normal speech, or at the other extreme to be emitted quickly and violently as in the case of a shout. Between these two extremes we must conceive of the rate of expulsion as a continuum covering all degrees of air escape from the lowest to the highest possible, ranging from the most economic use of the outflowing air current under strict muscular control to the quicker escape of air in a state of relaxation.

These physical processes provide the operational terms by means of which certain qualities of expression in speech, which may range from the monotonous and dry, through the warm and emotional, to the positively effusive and excited gush of speech, are becoming tangible and measurable. This assumption is supported by Stetson's (1928) findings resulting from his studies of speech movements in action. Stetson showed the

divisions of the syllables to be identical with the divisions of the chest pulses one from another by demonstrating experimentally that the syllable movement equals the 'breath pulse' which equals the 'chest pulse'.

Thus the syllable has been shown to be not only a phonetic unit but also a unit of respiratory activity, i.e. a motor unit of the respiratory apparatus. *The number of syllables uttered during each expiration would therefore measure the rate of expulsion or escape of the outflowing air in speech.*

In terms of the process measured, the more syllables spoken during an outlet of breath the slower the pace at which the inspiratory muscles pass from a state of contraction to that of relaxation, i.e. at a constant speed of talking. What is measured is therefore the delay of the relaxing phase of the respiratory cycle, effected by voluntary control, a motor phenomenon which seems analogous to Luria's (1932) delayed action as controlled by his 'functional barrier'.

The actual length of the contraction-relaxation phase depends on the rate of respiration as well as on the rate of speech. For the number of syllables uttered per outlet of breath tells us how far the energy supplied by the outflowing air current was under cortical control, i.e. how far it has been guided into the channels of speech activity, and how efficiently it has been utilized, i.e. into how many speech units it was transformed. Or, by reversing our viewpoint, we learn from it to what extent the air current returned free and unbridled, as a spontaneous and unexercised carrier of emotions, as in laughter, sighs, shouts, etc., and what proportion of it escaped per speech unit (degree of ventilation in speech).

We may therefore expect that the number of syllables per expiration, i.e. the degree of ventilation in speech, will prove to be a measure of the extent of the transformation of the energy of expiration into verbal activity, that is of the extent to which the involuntary action of respiratory reflexes is directed by the cortical action of producing speech sounds.

The testing of these hypotheses by means of the above measures is in progress.

VI. Summary

The aim of the investigation was to develop indicators of changes in emotional states based on the speech process itself. Previous research had suggested that respiratory behaviour in speech might be a fruitful line to study. A simple method of recording breath intake from sound recordings was discovered. Two measures were derived from speech-breathing activity: the rate of respiration during speech and the output of speech (number of syllables) per expiration. Both of them proved to be highly reliable and sensitive, but the output of speech per expiration was the more constant, as well as more sensitive, of the two. The phonetic and physiological facts of speech production and psychological evidence and theories, in respect of respiratory behaviour, suggested interpretations of the psychological significance of these measures which are in the process of being tested.

REFERENCES

ALTSCHULE, M. D. (1953). *Bodily Physiology in Mental and Emotional Disorders.* Chapter: Respiration. New York: Grune and Stratton.

AX, A. F. (1953). The physiological differentiation between fear and anger in humans. *Psychosom. Med.* XV, no. 5, 433–42.

CLAUSEN, J. (1951). *Acta psychiat., Kbh.*, Suppl. 68. Copenhagen: Ejnar Munksgaard.

FENICHEL, O. (1953). *The Collected Papers. First Series.* New York: W. W. Norton.

FREUD, S. (1936). *The Problem of Anxiety.* New York: Norton.

GOLDMAN-EISLER, FRIEDA (1952). Individual differences between interviewers and their effect on interviewees' conversational behaviour. *J. Ment. Sci.* XCVIII, 660–71.

GOLDMAN-EISLER, FRIEDA (1954*a*). A study of individual differences and of interaction in the behaviour of some aspects of language in interviews. *J. Ment. Sci.* C, 177–97.

GOLDMAN-EISLER, FRIEDA (1954*b*). On the variability of the speed of talking and on its relation to the length of utterances. *Brit. J. Psychol.* XLV, 94–107.

GOLLA, F. L. (1921). The objective study of neurosis. *Lancet,* II, 115–22.

HARNIK, J. (1930). One component of the fear of death in early infancy. *Int. J. Psycho-Anal.* XI, 485–99.

KUBIE, L. S. (1948). Instinct and homeostasis. *Psychosom. Med.* X, 15–30.

LURIA, A. R. (1932). *The Nature of Human Conflict.* New York: Liveright.

MEYER, G. H. (1892). *The Organs of Speech.* London: Kegan Paul Trench and Trübner.

MOWRER, O. H. (1953). *Psychotherapy, Theory and Research.* New York: Ronald.

RUESCH, J. & PRESTWOOD, A. R. (1949). Anxiety, its initiation, communication, and interpersonal management. *Arch. Neurol. Psychiat., Chicago,* LXII, 527–50.

SCHILLING, R. (1929). Über 'Inneres sprechen'. *Z. Psychol.* III, 204–55.

STETSON, R. H. (1928). Motor phonetics, a study of speech movement in action. *Arch. néerl. Phon. exp.* III, 1–216.

SUTER, J. (1912). Die Beziehung zwischen Aufmerksamkeit und Atmung. *Arch. ges. Psychol.* XXV, 78–150.

TRAVIS, L. E. (1931). *Speech Pathology.* New York and London: Appleton.

(*Manuscript received* 20 *July* 1954)

CORRESPONDENCE

The Editor
British Journal of Psychology

Dear Sir,

I would like to comment on the paper by Prof. Oldfield.[1] If I appear to write *ex cathedra* and without explanation of terms, the limitations inherent in a letter must be blamed.

The analogy of an inscription on stone is not inappropriate to the initial registration of events. Moreover, it is quite possible for such an inscription to wear away in such a manner that, at various stages, it appears to be in different 'scripts' and thus to have different meanings. This would be an example of 'disconnected ambiguity'[2]—an important phenomenon related to the dimensionality of the coding. The real objection to the 'inscription on stone' type of analogy is the implicit, but not necessarily true, assumption that it does *not* give rise to this kind of ambiguity during erasure. But disconnected ambiguity is a striking feature of perceiving and remembering, and I suggest the time has come when we can make these more fundamental concepts explicit.

It is true that the known functions of regenerative loops in the central nervous system 'relate to the temporal prolongation of inhibitory and excitatory effects'. The point is that these prolongations represent storage of information, and hence a crude but genuine form of memory. Probably the physiological techniques applied to such loops are incapable of detecting any form less crude.

The 'circulating memory' employed in computers does not, in any relevent sense, embody a 'quite novel principle' (contrary to the author's assertion). The essential principle is that of preserving the message by repeatedly rewriting it before it has become unintelligible, just as ancient documents may be copied or archeological relics photographed before they crumble into dust—or as we deliberately rehearse a verbal message in order to fix it more firmly. It is worth remembering that the ordinary strategems and artifices of mankind will provide nearly all the qualitative models we need, without recourse to special technologies. The use of binary number codes is not a technology but an unexceptionable logical device, so far as we are concerned.

The statement that 'economy can be effected (by the type of coding proposed) without impairing accuracy of reproduction' is only true if there is no chance of error anyway—i.e. if the system is 'noise-free', in the jargon of information theory. Otherwise, minimum-redundancy coding (redundancy is a measure of the superfluity of symbols in a message) must be paid for by a greater frequency of errors. The nervous system is very far from noise-free, and we ignore the fact at our peril.

For example, the 'sudden' recoding which, as the author suggests, manifests itself as 'insight' need not be left, as he leaves it, to some mysterious and *ad hoc* 'active searching'. Given the necessary conditions for disconnected ambiguity, it will come about sooner or later with the same kind of probabilistic inevitability as ensures the increase of entropy in a closed physical system.

[1] Oldfield, R. C. (1954). Memory mechanisms and the theory of schemata. *Brit. J. Psychol.* XLV, 14–23.
[2] Woodward, P. M. (1953). *Probability and Information Theory, with Applications to Radar.* London: Pergamon Press.

As a further *ad hoc* hypothesis, the author invokes a special feedback mechanism to minimize redundancy. Possibly he did not intend it as *sui generis* and distinct from the basic mechanism of learning. Be that as it may, a simple way of looking at the matter is to note that, since the greater the redundancy the greater the protection against 'noise', that form of coding which will longest withstand the attrition of such random disturbances is the one which attributes the least information to the original message. Therefore, given the necessary coding filter for successively lower levels of information (increasing the redundancy attributed to the original message), the tendency to minimize the storage space occupied is automatic.

Incidentally, diminishing the storage space occupied is not to be understood as reducing the number of symbols in the store, but as reducing their total entropy. When you clear the register of a desk computer, it contains exactly as many symbols as before—only they are all noughts.

The formation of the encoding filter (or scheme of classification, as it may more intelligibly be termed) can, at this level of abstraction, be best regarded as the familiar operation of averaging—i.e. accumulating measures of statistical frequency, which act as 'probabilities'. A frequency-count destroys (by ignoring) an enormous amount of information; consequently it would appear that the local entropy is reduced at the cost of a vast increase of overall entropy, as general principles require.

The elementary processes of attrition and accumulation, together with disconnected ambiguity, seem, therefore, to be capable of exorcizing most of Prof. Oldfield's ghosts.

Yours faithfully,

W. E. Hick

Psychological Laboratory
University of Cambridge
3 *March* 1954

PUBLICATIONS RECENTLY RECEIVED

The Unconscious Origins of Berkeley's Philosophy. By J. O. Wisdom. London: Hogarth Press. 1953. Pp. xii + 244. 25*s.* net.

The kernel of this book is an attempt to apply to Berkeley's philosophy a psychoanalytic interpretation. Wisdom uses current psychoanalytic theory of the British School to give a psychoanalysis of Berkeley and thereby present the psycho-genesis and function of his metaphysic. The central theme which emerges is Berkeley's infantile attitude to faeces, and how this led him to the doctrine of 'esse percipi'—a doctrine which cleansed the world for him of the external poison of 'matter'. I have no room here to present any of the details of Wisdom's analysis. It will have to be sufficient for me to say that Wisdom seems to have made a careful and interesting application of psychoanalysis, and one which analysts themselves will no doubt consider to be a contribution to the subject.

Now I shall pass over the objection that it is really not worth while attempting the psychoanalysis of the dead, since, in the absence of the effects of interpretations upon the patient, the whole enterprise becomes too speculative. I shall also pass over the general objection from the hard-headed members of the British Psychological Society, namely that the whole of Wisdom's analysis of Berkeley is just sheer phantasy anyway and not worth bothering about. What I shall do is to use this review to touch on one problem of importance for psychologists that this book raises but does not develop or answer. This is the problem of the relation between 'the causes of' and 'the reasons for' belief, between the psychology and the logic of thinking. Wisdom's book raises this problem in the following way.

Wisdom says that 'the chief aim of his book is to throw light on philosophy' by psychoanalysing Berkeley, the man and his philosophy. In other words, he offers us the key sources (or causes) of Berkeley's philosophical position, and claims that thereby he is doing something of philosophical relevance. However, I can only discover one paragraph (on page 230) in which he does try to tell us how his book 'throws light on philosophy', and unfortunately I find this paragraph so confusing that it fails to enlighten me. I suspect that other British philosphers will find themselves here in the same difficulty as myself. I think it is a pity that Wisdom did not go on to develop the philosophical relevance of psychoanalytical interpretations, and thereby show the bearing that psychological 'causes' have on 'the reasons' in which philosophers are interested. For, in my opinion, the analysts have got hold of something here which is well worth developing. A brief review is hardly the place to exhibit the relevance of these psychoanalytical interpretations, but I shall nevertheless try to indicate one way in which this may be done.

Consider Berkeley's arguments for the view that 'esse percipi'. When we come to understand this view, we also appreciate that, far from being false or fantastic, it is defensible and true—to put the point crudely. Yet most of us at present have no wish to talk like this, and contemporary philosophers would maintain that we are not logically obliged to do so either. But when Berkeley argued that 'esse percipi', he quite obviously thought that by his arguments he *had* provided 'a good reason' why we are logically obliged to talk like this. To-day we can appreciate the force of Berkeley's arguments, but we do not accept them as constituting a good reason for having to say that 'esse percipi'. Why, now, did Berkeley accept his arguments as constituting 'a good reason'? In answering this question we have to ask why this way of talking—namely, 'esse percipi'—resolved his own philosophical conflict. If Wisdom is right, part of the answer will be that this way of talking helped to resolve his philosophical difficulty about matter, and it did so because this conflict was, in part, the product of a psychological disorder which had its origins in his childlike attitude to faeces. To talk of 'esse percipi' satisfied him partly because it helped to resolve the psychological problem from which he suffered and which contributed to form the character of his philosophical difficulty. His implicit criterion of 'a good reason' in this context is determined in part by his psychological disorder and his early infantile attitude to faeces. If Wisdom is right, he has drawn attention to certain unconscious sources of Berkeley's philosophical difficulty and of his compulsion to resolve it in the way that he did.

We can now point to the philosophical relevance of psychoanalytic interpretations. They are relevant to the theory of philosophical activity, since a description of what philosophers are trying to do has to allow for the fact that they are trying to resolve philosophical difficulties which may sometimes be generated in part by sources of the type with which psychoanalysis has familiarized us. This suggests that there may not be one 'right' or 'correct' way for everyone of answering a philosophical question. It

suggests that the appraising terms of philosophical discourse, like 'good reason', 'right (or wrong) in saying this', 'mistake', 'true', 'false', 'correct', and so forth, need an elucidation that brings out their personality dependence. When, however, we appreciate the theoretical relevance of psychoanalytic interpretations, the manner in which we *practise* our philosophical activity is likely to be affected in various ways. Consider one such way. Suppose I point out—as a contemporary linguistic philosopher may do—that 'much of Berkeley's theory does not contradict common sense' (Warnock, *Berkeley*. Pelican Books, 1953). If I do *not* appreciate the theoretical relevance of psychoanalytic interpretations, I may easily imagine, as many contemporary philosophers appear to do, that this statement about Berkeley's theory is true in the same unqualified or universal way as is the statement that 'Everest is the world's highest mountain'. But if I come to appreciate the point of psychoanalytic interpretations, I am *less* likely to imagine that my statement about Berkeley's theory is true in the same straightforward, unqualified way as is my statement about Everest. For I can now appreciate that my statement about Berkeley, in contrast usually with the statement about Everest, may be part of an attempt to deal with a psychological conflict of my own, or with one fairly common at the present time in this country among the products of the Public Schools. It may be that what makes me want to emphasize this statement about the commonsensical character of Berkeley's theory and feel satisfied when I have shown its force is, in part, my own compulsive adherence to the ordinary speech learnt at my mother's knee; and this in turn may go back to an infantile wish to return to the ordinary milk and secure breast of my mother. It follows that for those who are *not* in a psychological conflict similar to mine, my statement about Berkeley may be uninteresting, pointless and not at all satisfying. To such people my statement about Berkeley will be true but not obligatory, since I have not given them 'a good reason' for talking in this way about Berkeley. Consequently, when I appreciate the point of psychoanalytic interpretations, I am likely to use appraising terms like 'good reason', 'mistake', and the like, much more cautiously and in a way which tries to make explicit my own criteria governing their use. This may have a further consequence. It is well known that much philosophical activity in England at present consists of a naturalistic exploration of our speech and language—a pursuit which is apt to degenerate into dictionary mongering and armchair social anthropology. When we appreciate the point of psychoanalytic interpretations, we may cease making this pursuit the central, or dominant, feature of our philosophical activity. We may be helped to recall that philosophical activity is only worth while if it is interesting, and that it is only interesting if it deals with philosophical *difficulties* which *really* trouble us.

But of course, to say all this is merely *to touch* on the philosophical relevance of psychoanalytical interpretations. I have not even pointed to their obvious limitations, namely that they do not deal with those unconscious sources of our philosophical difficulties which are to be found in our verbal habits—sources which the work of Wittgenstein and John Wisdom have emphasized, and which, unlike those psychoanalysts point to, have an immediate and critical relevance for philosophical activity. It is because this whole theme is exciting and obviously important that I am sorry Wisdom did not take the opportunity provided by this book to give us his views on the philosophical point of presenting us with the unconscious origins of Berkeley's philosophy. If he had done so, he would certainly have helped to clear up the current confusion in theoretical psychology over the relation between the logic and the psychology of thinking and belief.

B. A. FARRELL

Research Methods in the Behavioral Sciences. Edited by L. Festinger and D. Katz. London: Staples Press Ltd. 1954. Pp. xi+660. 50*s*.

Books on research methods tend to be rather tedious, particularly in fields where there is little agreement on methodology. Books written by a large number of different authors tend to be repetitive, lacking in uniformity of style, and evincing no clear-cut plan or design. *Research Methods in the Behavioral Sciences* is singularly free from both these faults. It is well planned and put together, the individual papers complement each other without overlapping, the style of writing is uniformly good throughout, and the contributors almost without exception have written at a very high level of professional competence indeed. The book is likely to become the standard text in this field for many years to come.

One of the reasons for this unusual success derives, undoubtedly, from the fact that all those concerned in this production are associated with one university. The University of Michigan, through its Institute of Social Research, which combines the divisions of the Social Research Centre and the Research Centre for Group Dynamics, was in an unusually favourable position for undertaking a task such as this, due to the exceptionally large number of competent and experienced social psychologists, sociologists, and other

social scientists working there. Much of the material reviewed in this book consists of original research, frequently unpublished, of members of the Institute; this intimate acquaintance with, and appreciation of, the complexities of social research is clearly reflected in the very sophisticated discussion of the problems which arise. It is difficult to single out chapters for special praise. To the reviewer, Helen Peak's 'Problems of Objective Observation' and Clyde Coombs's 'Theory and Methods of Social Measurement' appear particularly outstanding, the former giving the best discussion to date of different types of measuring and scaling devices, the latter putting forward a new, and in many ways, highly original theory of psychological measurement.

For the serious student, this book will certainly become required reading, and even the expert will learn a good deal from it. To British readers who may not always receive the duplicated or printed research reports so extensively quoted in this book, it may in addition serve as a reminder of the many different areas in which social psychologists in the United States are at present carrying out research. They may also feel that the almost complete absence of such research from the British scene does not reflect favourably on the prestige of British Psychology.

H. J. EYSENCK

Experimental Psychology (revised edition). By R. S. Woodworth and H. Schlosberg. New York: Holt and Co. 1954. London: Methuen and Co., Ltd. 1955. Pp. xii + 948. 63*s*.

It is sixteen years (1938) since the first edition of Woodworth's *Experimental Psychology* was published, and every psychologist who has used, read, or referred to this valuable survey of laboratory psychology will rejoice that a second edition is now available and that Prof. Woodworth himself has been able to play a part in its preparation. The new edition has about the same number of pages as the old, but the use of double columns has enabled about 20% more matter to be squeezed into it. A great deal of new material has been included. In fact, nearly half the references are to the work of the last 15 years and over 500 of them to books and articles published since 1950. Prof. Woodworth and his co-author, Prof. Harold Schlosberg of Brown University, have done a good job of revision.

Unfortunately some material from the older edition has had to be omitted. The chapters on feeling and on experimental aesthetics have gone altogether: accounts of older work have been curtailed. The amount of space, for example, devoted to the older work on memory has been considerably reduced. Hence, as the authors are at pains to point out in their introduction, the new edition does not altogether supersede the old, which still remains valuable for reference. The principal additions, apart from the fresh material added to every chapter, are an entirely new chapter on scaling methods and a greatly increased treatment of animal learning. The inclusion of the latter was probably inevitable if the book is to serve the needs of American classes, but it is not full enough to be really useful except to the undergraduate student, and it duplicates a great many other surveys of learning and learning theories. Since the authors have endeavoured to retain the 'style, viewpoint and general level of difficulty' of the first edition, a case could have been made out for excluding animal experiments, which were only very briefly mentioned in the earlier volume.

In general, however, Woodworth and Schlosberg have succeeded in their aim. This is still recognizably the same Woodworth in a rejuvenated form. Perhaps it seems on the whole rather more elementary in its treatment than the older edition, due no doubt to the greatly increased material which has been compressed into not very much greater space, and the necessarily rather cursory treatment of many topics, where progress in knowledge has been rapid. And possibly the eclectic approach which it adopts seems somewhat less satisfying to a generation that has become much more interested in and impressed by the importance of theory. Still the stuff of experimental psychology is here—here are accounts of a great many important recent investigations, by Michotte, Ames, Gibson, Sherif and dozens of other contemporary workers.

L. S. HEARNSHAW

L'Année psychologique. 53ème année, 1953. Paris: Presses Universitaires de France. 1500 fr.

British psychologists are inclined to feel a little indignant when their American colleagues cannot spare the time to sample the good things which, for all they know, abound in our publications. Yet our own attitude to Continental psychology is remarkably similar. *L'Année psychologique* is a convenient means of reducing our ignorance, at least as far as France, probably the most psychologically active of our near

neighbours, is concerned. The latest issue follows the usual pattern: we have some original articles, some reprints of lectures delivered before the Association de Psychologie Scientifique de Langue Française and summaries of psychological articles and books that have appeared recently.

The summaries show not only that French psychologists are diligent in scanning British and American products but also that they are better placed than ourselves for observing developments in Central Europe. But the topics selected and the space allotted to them are strikingly different from what we should expect in comparable English language reviews and reminiscent of what they would have contained some decades ago. 'General psychology' is granted less than a quarter of the pages in question. There is much discussion of detailed analyses of sensory and motor processes but little of the central processes that link the two, apart from the contributions of neurophysiologists and psychoanalysts. The areas that most British and American general psychologists would consider of most moment—learning, motivation, perception, symbolic functions, theory-construction—receive little or no mention. There is little evidence of interest in the various efforts that have been made, in the U.S.A. and other English-speaking countries, to use systematic theoretical models to integrate the various fields of psychology and to guide research. Piaget's paper, in which he points out some parallels between his recent thinking and that of the cyberneticians and the Chicago mathematical biophysicists, is a thought-provoking exception. Reuchlin contributes a valuable article on research design, but he writes in a way suggesting that the topic is new and not accepted without suspicion in France.

Highly stimulating reviews of recent work in neurophysiology are provided by Piéron and Ajuriguerra. We can see from them what the French psychophysiological tradition must still command our respect. But they also make it clear to us once again that we cannot afford to wait for the neurophysiologist to solve all our problems for us and that the faster we progress in devising scientific psychological techniques for working out the laws of behaviour the better it will be for both of us.

This volume, in short, is one more reminder that psychology is still not an international enterprise like other sciences but that it is still all too neatly parcelled up into national allotments. But we can help to remedy this by adopting the habit of looking at *L'Anneé psychologique* and at the same time find much that is well worth watching in French psychology.

D. BERLYNE

The Design and Analysis of Experiment. By M. H. Quenouille. London: Charles Griffin and Co. Ltd. 1953. Pp. xiii+356. 36*s*. net.

In recent years at least half a dozen texts have appeared each presenting, at varied levels of sophistication, techniques suitable for the design and analysis of certain types of experiment. In view of their almost exclusive devotion to the methods of analysis of variance it is unfortunate that the titles of these works, like the one under review, tend to suggest that these methods are applicable to the analysis of data obtained from any type of experiment.

The first half of the text under review gives an account of methods suitable for analysing of data obtained from experiments designed in the form of randomized blocks, Latin squares, factorial schemes, and incomplete blocks. These are topics which may be found in most books on the subject but readers with no previous acquaintance of statistics will find difficulty in following the rough intuitive arguments that are given as justification for the methods of analysis. From this point of view the work suffers in comparison with the well-known work by Snedecor.

The second half of the book breaks new ground. In particular, the author discusses methods by which we may estimate the residual effects consequent upon applying a sequence of treatments to the same subject. For the experimental psychologist, what can be done in this direction still appears to be rather limited, but the suggested techniques will be of great interest to all who are persuaded that analysis of variance procedures are invalidated if repeated observations are made on the same subject. Other chapters deal with techniques for combining the results obtained from several experiments, the appropriate analysis to be employed when observations are missing, and the various scaling methods available for transforming data into a form suitable for the application of analysis of variance.

The expert will find no clear account of the fundamental probability models underlying the various types of analysis nor any details of the least squares approach to the subject. He will, on the other hand, appreciate the discussions devoted to the more complex experimental designs which have been recently developed and which, up to now, could only be found in journals. Further, he will find that the examples given to illustrate the methods, though remote from psychology and drawn mainly from the agricultural and medical sciences, are discussed with commendable common sense.

In general the book appears to be suitable neither for the beginner nor the expert, but is invaluable for the treatment it gives of many specialized topics related to analysis of variance which are not to be found in existing texts.

A. R. JONCKHEERE

Traité de psychologie animale. By F. J. J. Buytendijk. Paris: Presses Universitaires de France. 1952. Pp. xv + 362. 1000 frs.

The title of this work is misleading; it cannot by any reasonable criterion be considered as a 'Treatise' on animal psychology. The author himself makes this clear in the introduction, remarking that he has omitted any treatment of tropisms, the behaviour of unicellular organisms, perception, and learning phenomena. To most readers it will seem as if some of the main topics in animal psychology have been ignored. But Prof. Buytendijk has other aims in mind than merely giving a systematic account of the subject. By means of discussions of topics such as laughter and play in animals and whether chimpanzees can 'lie', the author attempts to introduce a new approach to the study of animal behaviour. This new approach is unfortunately a type of phenomenological analysis popular in literary and philosophical circles on the Continent and which has now, with the work under review, reached the frontiers of animal psychology. The result is that a distinguished authority in the field has produced a book which by the opacity of its style (though it should be remembered that we have here a translation) and the looseness of its thought, will dismay readers who are acquainted with the author's earlier empirical work. The book may be recommended, however, as an example of one of the prevailing trends in contemporary Continental psychology.

A. R. JONCKHEERE

Physiologische Psychologie der Sprachvorgänge. By Friedrich Kainz. Stuttgart: Ferdinand Enke Verlag. 1954. Pp. 571. Paper, DM. 43.50; linen, DM. 47.00.

Prof. Kainz's book on the physiological psychology of speech continues a traditional line of Viennese psychology. It is, after Bühler's *Sprachtheorie*, the most important work in a field which has been much neglected by psychology at large. Between the recent revival of interest in the study of language behaviour and Wundt's emphasis on the utmost psychological importance of the phenomena of language lies a fallow period during which the study of language was neglected by psychologists and left entirely to those interested in linguistics, phonetics, and the physiology and neurophysiology of speech. An enormous body of basic knowledge has grown out of the activities in these fields, of which the general psychologist is largely unaware. With the growing interest in the subject of communication language behaviour demands more and more attention as an area of investigation.

Prof. Kainz's book is, at this juncture, invaluable to any student or worker in the field of linguistic psychology. It is marked by unusual breadth, is encyclopaedic in scope and immensely serviceable in as far as it puts before the reader a wealth of well-digested material. If, under these circumstances, the book has grown to awe-inspiring proportions, the advantage of the unstinted supply of information given is that it allows the reader to form his own opinion in cases of controversy, even when Prof. Kainz makes his viewpoint clear. No advantage, however, seems to accrue from the long-winded and long-worded style, even by German standards, which one fears will deter many a potential foreign reader with a working knowledge of German.

The volume reviewed here is the third of the four volumes comprising the whole work. In the two preceding volumes Prof. Kainz dealt with the properties and achievements as well as with the origin and development of language. In the volume under review he surveys all those peripheral and central processes which come into operation in the act of speaking or of understanding spoken language. Prof. Kainz sees his task to include the presenting of a comprehensive psychology of speech based on the physiology of speech processes, as well as to give an historical survey of the science and the problems of linguistic psychology.

The book is divided into five chapters. In the introduction the author defines the science of linguistic psychology as dealing with language as an emotional, mental and physical activity of man. Its subject is the totality of the psychophysical processes and functions which operate in the production of speech. Prof. Kainz distinguishes two processes: (i) the central or cortical activity of diction, and (ii) the peripheral process of articulation. Speech-physiology therefore deals with the processes in the cortex and the brain stem, the intra-cerebral paths participating in the speech activity, and the afferent and efferent pathways in the nervous system activating the muscles of the organs of respiration, phonation, and articulation.

The psychological aspect of speech lies, according to Prof. Kainz, in the conscious intention, the verbal imagery (acoustic and kinesthetic), the schemata of sentences and beginnings of phrases, which are completed according to the initial plan, and finally the control and guidance of articulation.

The second chapter deals with the problem of localization and attempts to answer the question of what are the processes activated in the cerebral cortex during speech activity. The author's position in this controversy is a compromise. Prof. Kainz prefers to remain within the principle of localization critically modified in the light of Gestalt view as represented particularly by Goldstein.

Chapter three deals with the activity of speech. We are taken through the detailed phases of the speech process, from the initial voluntary impulse to speak to the detailed movements of articulatory execution which follow automatically. Prof. Kainz makes a clear distinction between the phases of thinking, diction or 'inner speech', and utterance. Diction is not the behaviourist 'sub-vocal talking' equivalent to thinking, but the totality of processes and dispositions which is responsible for the formation of verbal symbols, and their grammatical and syntactical structuring.

This precision of analysis in pursuing the actualities behind each term is characteristic of the whole book. Such ground-work enables the experimental worker to formulate concrete and testable hypotheses.

Of particular interest to the experimental worker in the field of speech behaviour and communication are those parts of this chapter which deal with the act of speaking. The discussion includes the action of the peripheral organs of speech which take part in the three processes basic to speech production: the respiratory organs, the organs of voice production, and those of articulation. Prof. Kainz then discusses the co-ordination of speech and hearing, the organic basis of articulation, and the relation between psychology and phonetics. In this chapter alone more than 350 references are cited. It is a valuable experience for any psychologist whose work takes him into the fields of interviewing or psychotherapy, for example, to be made to realize how many aspects there are to the basic activity of speaking alone and how much there is to be known about it.

While all these need to be considered in any comprehensive science of communication it is well to take heed of Prof. Kainz's warning that the study of these phenomena can only become fruitful from the point of view of psychology if they are brought into relation with the consciousness of the individual speaker. Such emphasis is the more salutary when seen in juxtaposition to the fascination exercised on psychologists by the achievements of communication engineers and information theory.

A further chapter is devoted to the subject of the understanding ('verstehen') of speech. In the last chapter Prof. Kainz deals with the subject of conversation. Consistent with his general outlook which has emerged as one of dynamic functionalism moderated by a strictly biological orientation, he rejects the behaviourist theories of conversational speech as being a mere mode of social action. He insists that expression cannot be limited to the monologue but that it is one of the functions of conversation. Through conversation, Prof. Kainz impresses upon us, we can study linguistic activity in the fullest realization of its potential powers, as a concrete linguistic event containing and integrating all functions of speech, active and receptive.

Psychiatrists and interviewers who are using conversation as the vehicle and instrument for their work will welcome this conclusion, for they will feel that it carries conviction in terms of their own experience. At the same time Prof. Kainz has provided a solid background to the exact study of the linguistic process by showing into how many component processes the process of speech can be broken down.

While these facts might have been presented less turgidly, and the index system arranged better to suit the reader's convenience, these are minor points if measured against the merit of the book. Prof. Kainz's service to the scientific study of psychological phenomena based on the use of speech and of linguistic psychology cannot be emphasized enough.

F. GOLDMAN-EISLER

Communication and Persuasion. By Carl I. Hovland, Irving L. Janis and Harold H. Kelley. Yale University Press; London: Geoffrey Cumberlege. 1953. Pp. xii + 315. 36*s.* net.

Like all good things, the present trend towards greater scientific rigour in psychology can be carried too far. There is a growing tendency to feel that even the most incontestable statement about human behaviour and motives must be supported by statistical evidence, and a good deal of time and money is being spent on the experimental demonstration of the obvious.

A few such demonstrations are included in the book under review, which reports the results of a co-ordinated research programme conducted at Yale University on 'the modification of attitudes and beliefs

through communication'. Thus one experiment demonstrates that if we greatly value our membership of a particular group we are likely to be resistant to views which conflict with the group's standards; several others, that we are more likely to agree with a statement, and to judge it as 'fair' and 'honest', if it comes (or is believed to come) from a source we regard as reliable than if it comes from a source we suspect.

But it would be unfair to judge the book by these examples; not all the results reported are so predictable *a priori*, and some are of considerable practical interest. For example: a fear-arousing, gruesomely illustrated lecture on dental hygiene produced significantly more emotional tension than a less fear-arousing lecture, but had significantly less effect on the subjects' future behaviour; one-sided advocacy of a particular view produced less resistance to counter-propaganda than did a statement which took account of the arguments for the other side; students developed a more favourable attitude towards a point of view after they had carried out the assignment of presenting it as favourably as possible to an audience; the most 'persuadable' students tended to be shy, unaggressive and liable to depression, while the least 'persuadable' showed tendencies towards neurotic anxiety and obsessional symptoms.

Besides describing the experiments carried out by the research team, the book contains valuable summaries of other relevant work. Though the style is as graceless as the current American convention demands (e.g. 'Fear of social disapproval might give rise to strong facilitating motivations with respect to acceptance of persuasive communications') the book is by no means unreadable; and it is admirably produced and printed.

M. KNIGHT

The Study of Language. By JOHN B. CARROLL. Harvard University Press; London: Geoffrey Cumberlege. 1953. Pp. xi+289. 38*s.* net.

This is a specialized book and will appeal mainly to those with a professional interest in linguistics, but many psychologists will be interested in the chapter on linguistics and psychology. The author starts from the axiom that thought differs 'only in the matter of degree of complexity' (i.e. only in complexity?) from simpler response-processes such as pressing a key in response to a signal; and he makes a provisional attempt to fit 'language behaviour' into the framework of S–R and reinforcement-theory. To the present reviewer it does not seem a promising approach; but some of the younger generation may feel stimulated and enlightened to learn that when they talk they are making socially reinforced verbal responses, and that when they use verbally expressed hypotheses as an aid to problem-solving 'the selection of linguistically-coded hypotheses is a function of two factors: (*a*) the stock of verbal responses which the subject can call up to solve the problem, and (*b*) the relative strengths of these responses in relation to the stimulus values of the material presented'.

M. KNIGHT

Developments in the Rorschach Technique. Vol. I. *Technique and Theory.* By BRUNO KLOPFER, MARY D. AINSWORTH, WALTER G. KLOPFER and ROBERT R. HOLT. London: G. G. Harrap and Co. Ltd. 1954. Pp. xi+726. 42*s.*

A full appraisal of the eagerly awaited 'new Klopfer' must wait until both volumes of the work are available. Meantime, the percipient reader will note that the present volume, by itself, is nearly twice the size of the familiar 'Klopfer and Kelley' of 1942. This comparison not only indicates the immense amount of Rorschach research that has been carried out in these twelve years; it also reflects the increased care and clarity with which most of the basic concepts are set out, and the much fuller theoretical discussion now provided (Part Three, entitled 'Theory', extends to nearly 200 pages). Of probably greater value still is the series of explicitly stated 'interpretative hypotheses' introduced at the outset of the discussion of each scoring category or relationship among factors. The effect, for the beginner, is to make the interpretative significance of the various types of response both clearer and more convincing; for the experienced Rorschach worker it affords an opportunity of re-examining the theoretical bases of the technique, which he might otherwise have been liable to take for granted.

Certain changes in practice, foreshadowed in the 1946 *Supplement*, are now fully developed. Chief among these is the discarding of the $F+$ category in favour of 'Form-level rating', a device which now assumes major importance, but which to the present reviewer has always appeared clumsy and irrational, especially when the ratings, which are made on a scale which is discontinuous, are averaged. The whole question of quantification in Rorschach is one which still requires to be thrashed out; thus, percentages and 'normal limits' are frequently quoted, but almost inevitably with the proviso that below a certain

minimum (e.g. of total length of record) these findings cannot be applied, a safeguard which cannot but cast doubt on the validity of the relevant interpretation *within* the permitted limits.

Other innovations in method include the introduction of 'Sequence analysis', and of an 'Analogy period' (between the Inquiry and the Testing-the-Limits procedure). The former is almost self-explanatory in its intention, and strikes one as a valuable tool but one difficult to apply without undue expenditure of time. The conduct and function of the latter are, unfortunately, poorly defined. Additional contents include a very fully reported case-study (nearly 80 pages), and a description of a 'Prognostic Rating Scale' designed to predict a patient's response to psychotherapy.

It is perhaps a pity that the new book could not have been so written as entirely to supersede the earlier work. Much of the exposition (particularly in the first chapter of all, on 'Administration') is couched in language (not to say jargon) which presupposes familiarity with Rorschach concepts, and certain examples and other material previously cited are not repeated. That being so, one is tempted to look for redundancies, but in practically every case one cares to test the new is an improvement on the old. Consequently, the present work becomes one's vade-mecum, with Klopfer and Kelley retaining its place on the shelf, for reference.

BORIS SEMEONOFF

The Graphomotor Projection Technique. By S. A. Kutash and R. H. Gehl. Springfield, Ill., U.S.A.: Charles C. Thomas; Oxford, England: Blackwell Scientific Publications Ltd. 1954. Pp. xi+133. 27*s.*

In this test the subject is blindfolded and asked (in effect) to scribble freely, without trying to make anything in particular, for five minutes. This is repeated, after which there is an inquiry asking for recall of what has been done on the first attempt, the outlining of objects, etc., seen in the production and finally associations to each of the things seen or imagined. The inquiry is repeated for the second production.

The structural and other qualities of the production are classified according to carefully worked out criteria, and these features are related to the behaviour and inquiry data to build up a diagnostic and descriptive personality picture. A research is reported in some detail showing a large number of significant differences in the performances of normal and schizophrenic subjects and examples of normal and schizophrenic records and interpretations are provided.

Some claim is made for this projective technique that it is an attempt to realize an aim, which arose out of the Allport and Vernon Studies in Expressive Movement, that it is important to study movements not in isolation nor as single traits but rather in respect to the basic motivation of the individual. Although the authors for the most part adopt a holistic and dynamic approach, this claim is not substantiated. In the reviewer's opinion a projective technique designed for clinical use should be capable of throwing light on the nature of personality dynamics, conflicts and their causation in terms of concepts directly relatable to therapeutic work. The rationale of the test and the results demonstrated are in far too general terms for this purpose. As an aside one would also express doubts as to whether the test, in the way it is given, represents a task which patients can readily and meaningfully accept.

The book makes some thoughtful contributions to an interesting area of research; it may well be that it would have been more advisable at this stage to present the work as such, rather than as a clinical technique.

As with all books from this publishing company the production is serviceable and pleasing.

H. PHILLIPSON

Dynamic Psychiatry. Edited by Franz Alexander, M.D., and Helen Ross. Chicago: The University of Chicago Press; London: Cambridge University Press. 1952. Pp. xii+578. 75*s.* net.

This volume offers 'to students of psychiatry a comprehensive view of dynamic psychiatry'. The editors believe that the dynamic trend, essentially 'the advancement of the study of psychiatry from a descriptive into an explanatory phase', is the result of the impact of psychoanalysis upon the whole of psychiatry. This book, however, is not an exposition of psychoanalysis.

The book is in three parts: Part I, Concepts of Dynamic Psychiatry; Part II, Clinical Psychiatry; and Part III, Influence of Psychoanalysis on Allied Fields.

Part I consists of four chapters: I, Development of the Fundamental Concepts of Psychoanalysis, by Franz Alexander; II, Dreams and Rational Behaviour, by Thomas M. French; III, History of Metapsychological Concepts, by Edoardo Weiss; IV, Personality Development, by Therese Benedek.

Chapter II is disappointing. It is adapted from the author's *The Integration of Behaviour*, but is too brief to stand as the major section on dreams in such a volume as this. Chapters I and III should be read together. Weiss writes briefly on the early history of the concept of the unconscious and traces changes in metapsychological concepts. The writing is a little overladen with psychoanalytic terminology. There is not the contrast in these chapters which one might expect from the titles. Since both authors write historically or developmentally, a joint chapter might have made better exposition.

To the psychologist the title of Benedek's chapter is somewhat misleading. Perhaps *Psychosexual development of personality* would give a better guide to its content. Benedek applies the concepts in the chapters by Alexander and by Weiss to the developmental phases from infancy to senescence. For Benedek 'the development of the personality is the unfolding of an innate *anlage*—constitution under the influence of the environment'. This is an excellent chapter, but the argument would be easier to follow with sub-headings.

Part II: V, Neuroses, Behaviour Disorders and Perversions, by F. Alexander and L. B. Shapiro; VI, Acute Neurotic Reactions, by Leon Saul and J. W. Lyons; VII, Emotional Disorders of Childhood, by Margaret W. Gerard; VIII, Contributions of Psychoanalysis to the Study of Organic Cerebral Disorders, by H. W. Brosin; IX, Psychodynamic Approach to the Study of Psychoses, by J. C. Whitehorn; X, Contributions of Psychoanalysis to the Study of the Psychoses, by H. W. Brosin; XI, Principles of Psychiatric Treatment, by M. Levine. Chapter VI is mainly concerned with the so-called 'war neuroses'. The literature is briefly reviewed: Freud, Ferenczi, Abraham, Jones, Simmel, Kardiner, Radó, Grinker and Spiegel; incidence, etiology, types of stress, symptoms, dynamics of symptom formation, prognosis and treatment are among the topics dealt with.

Chapter V is perhaps the most disappointing in the book. One would expect to find here the full richness of the dynamic viewpoint. Instead there are 22 pages to cover definition of neurosis, hysterical conditions, obsessive compulsive states, depressions, hypochondriasis, behaviour disorders, alcoholism, drug addiction and sexual perversions. There are only eight references, and Fenichel is not included. Chapter VII should, perhaps, be read after Benedek's chapter. Under etiology there is a welcome statement of complex causation, 'the way in which a child reacts to any trauma is determined by the total constellation of constitutional trend, his status of physical maturation, his integrative capacity, his previous experiences, and the type of trauma'. The problems are arranged according to developmental levels. Treatment, diagnosis and prevention receive brief mention. This is a well written and concise account. There is no reference to recent work on autistic children, nor to the emotional problems arising on a secondary basis, e.g. as an accompaniment to orthopaedic handicaps, rheumatic fever, tuberculosis, etc. This is also a deficit in the book as a whole.

Brosin's chapters VIII and X are scholarly surveys of the literature and are not limited only to psychoanalytic references. In chapter X he hopes 'for a much more intensive research-therapeutic programme which will justify current assumptions that a large causative component of the psychotic process is psychological'. It is a brief chapter but covers 139 references from the Burghölzli group (1902–14) to B. Lewin (1950) on elation, and Eissler on schizophrenia (1951).

Whitehorn's chapter IX and Brosin's X should be read as a unit. Whitehorn writes, 'a psychosis is a form of reaction, not a mere deficit or impairment'. This chapter is weakened by the argument from psychotherapy. Levine's chapter XI needs evaluation by the practising psychiatrist to whom it is addressed, but to the reviewer it appears excellent. Broad in outlook, blending the goal of the scientific with that primary goal of being helpful to the patient, it discusses treatment at various levels against a frame of reference based on six considerations: clinical diagnosis; dynamic diagnosis, genetic diagnosis, transference, counter-transference and treatment possibilities. He briefly classifies techniques of psychiatric treatment and details the various types of psycho-therapeutic procedures. He relates this to the therapist's training and skill; 'only an accredited psychoanalyst should do psychoanalysis'. Clinical psychologists have therapeutic functions, 'if his training is adequate [he] may do the same varieties of psychotherapy [as the P.S.W., i.e., with close supervision by psychiatrists, using some of the aspects of the techniques of supportive relationship and expressive therapy limited to conscious material], while working as part of a psychiatric clinic team, although his primary responsibility is in the field of teaching, research, and the use of special diagnostic tests.' This chapter is especially recommended to the D.P.M. student and the clinical psychologist.

Part III: XII, The Psychosomatic Approach in Medicine, by F. Alexander and T. S. Szasz; XIII, Some Relationships between Social Anthropology and Psychiatry, by Margaret Mead; XIV, Clinical Psychology, by D. Shakow; XV, Animal Psychology and its Relation to Psychiatry, by D. M. Levy; XVI, A Review of the Influence of Psychoanalysis on Current Thought, by H. W. Brosin. The chapter by Alexander and

Szasz might profitably be read following chapter x. Wisely the authors avoid the reified term and write of the psychosomatic approach. They briefly review changes in concepts and discuss differences in present views, including that of the Chicago Institute. Application follows of these concepts to the problem of gastro-intestinal disorders, bronchial asthma, etc.

Chapter xiii by Margaret Mead contains no surprises for those who know her work; it is a thought-provoking survey along with speculation on her own research and that of others. Her final sections are on the problem of normality-abnormality and comment critically on Kardiner and Roheim. Though there is a reference to the work of Rivers with the Todas, and to Hose and McDougall, there is no reference to the 1924 papers of C. G. Seligman and Ernest Jones.

Shakow's chapter xiv is somewhat limited, perhaps because he defines clinical psychology as concerned 'with the psychological adjustment problems of the individual'. One might expect reference to experimental procedures, e.g. Ebbinghaus's saving method in relation to psychiatric problems such as Korsakow syndrome. He also follows the usual American error of dating clinical psychology with references only to Witmer and Healy. There are, however, valuable discussions of what questions (in the diagnostic realm) should be put to the psychologist to which he may legitimately be expected to have a reply. Testing is discussed in terms of controls set up to reduce the amount of dependence placed on the observer. Shakow points out how the unified psychological portrait of the individual patient depends upon sampling of psychological functions. His plea is for scientific sophistication in this sampling.

Levy's chapter xv argues the thesis that, having accepted 'our kinship with the animal world structurally and biochemically', we should not 'remain isolationist psychologically'.

Brosin's final chapter, xvi, supported by 150 references, refers to the numerous precursors of Freud, the growth of psychoanalytic influence on psychiatry in terms of persons, books and organizations, followed by summary references to its influences in psychiatry, sociology, anthropology, economics, religion and literature.

Asked to choose the ten best books in clinical psychology and psychiatry published in the last 10 years, the reviewer would have little hesitation in including this volume. The D.P.M. student will find it valuable and not likely to go unused, if only as a bibliographic source, when he is a consultant. The clinical psychologist should be familiar with much of the theoretical chapters but should profit considerably from part II and the chapter by Alexander and Szasz. The index contains no reference to Heredity. There are several saving references scattered in the chapters to hereditary constitution, to universal inheritance, and some mention of individual variation, but genetics on the whole is ignored. There is no mention of the evidence for inheritance in manic depressive psychosis and schizophrenia. The psychologist is likely to feel that the volume would be better for drawing on developmental psychology and the psychology of individual differences, and for more adequate use of the concepts of dispersion and the significance of differences. There is no discussion and little reference to criticism of the dynamic hypotheses. Some personal expectations of the volume were not fulfilled; there is nothing on suicide, experimental psychosis or ecstasy states.

J. C. KENNA

Group Dynamics: Research and Theory. By Dorwin Cartwright and Alvin Zander. London: Tavistock Publications. 1954. Pp. xiii + 642. 35*s*.

This is an important, comprehensive, and in places difficult, book. There are thirty-five well-chosen papers, and fifty-one contributors. It represents a well-integrated and frontal attack on the problem of application of scientific method to the study of group behaviour. While much of the book is informative and relatively easy reading, in places it is quite obviously the work of theorists, writing for other theorists, in a sophisticated and unavoidably technical way. As a contribution to psychological science the book has been greatly strengthened by the editors' decision to exclude papers which had 'no clear relation to any explicit theoretical conceptualization'. The research papers are classified in six sections, each section being introduced by a chapter written by the editors. The sections comprise: (1) Theoretical; (2) Group Cohesiveness; (3) Group Pressures and Standards; (4) Group Goals and Locomotion; (5) Group Structure; and (6) Leadership. A reasonably wide range of theoretical orientations is represented. Thus the first section includes papers by five theorists representative of somewhat different approaches. They include Cattell writing of group syntality from a factorial standpoint; Jennings as representative of Moreno's Sociometry; and Scheidlinger whose approach is psychoanalytic. Nevertheless, one would anticipate, in a book of this kind, considerable influence of the theory, concepts and researches of Kurt Lewin. This is broadly true, though only one paper by Lewin himself is included: a report on his classical studies of the effect of group decision on the changing of food habits. The editorial chapters are written from a standpoint

which might be described as a *development* of that of Lewin. The italicized word is important, since the book as a whole is a record of the considerable development that has occurred in recent years not merely of research, but also on the methodology side. If the editorial standpoint is 'Lewinian', it derives from Lewin in his later life: the *Field Theory in Social Science* period, not of the early 'topology' period. An opportunity appears to have been missed for taking account of major advances in recent years, in the understanding of the learning process. It is, for example, difficult to understand how the editors, in chapter 7, can discuss the increase and decrease of attractiveness (valence), of a group for its members, without reference either to the principles which govern reinforcement, or the process of extinction. Again, Lewin himself in his later years achieved close integration with the thought of the Neo-Gestalt wing of students of learning. In view of this it is disappointing that even Tolman finds no mention in the 642 pages of the book, other than (chapter 16), as a place name within the Massachusetts Institute of Technology.

As more detailed illustration of the content of this book we may instance the well-written editorial chapters on the six areas of 'group dynamics' examined; the interesting account by Schachter of the neglected problem of committee behaviour; several important chapters by Festinger and others on communication; investigations reported by Horwitz on the Zeigarnik effect in the field of remembering; and the intensive study by Gordon of a small group in terms of a single attitude, and its modes of expression. It is fortunate that a British edition of this book is now available and, in terms of its length, it is excellent value. There is much that can be read with profit by the student or intelligent layman, who can skip the difficult parts which the professional psychologist will have to 'work through' rather than merely 'read'. The volume as a whole contains many interesting and significant empirical observations. It represents also an account of theoretical and methodological advances which British Psychology as a whole cannot afford to ignore.

PETER MCKELLAR

Proceedings of the Association for Research in Nervous and Mental Disease. (December 14 and 15, 1951.) London: Baillière, Tindall and Cox. 1953. Pp. xiv+451. 68*s.* 6*d.* net.

This latest symposium, issued by the Association for Research in Nervous and Mental Disease, is now published in Great Britain.

The work is arranged in seven sections. Each section comprises four or five papers by authoritative speakers, followed by a discussion, and a presidential criticism and summary of the several contributions to the sectional theme.

Four sections are devoted to individual and group psychotherapy; two sections concern biological, electrical and pharmacological treatments, and two sections are given to neurosurgery.

The sections on psychotherapy combine recent advances with an historical account of almost every psychotherapeutic method and its field of application. This material is presented in a very readable and digestible form. No one technique is given prominence over others. Each technique, be it long-term or short-term treatment, individual or group treatment, hypnosis, psychosomatics, environmental manipulation, etc., is described with emphasis on its range of application and its limitations. Here one finds no mere account of the birth pangs of psychotherapy. The subject is presented as a mature science in which formerly conflicting theories and warring 'schools' have now developed into essential pieces of the total pattern. Thus, the student who is in doubt as to the range of application of various psychotherapeutic methods will find this book a really practical help.

The sections which deal with insulin treatment, electrical treatment and various pharmacological treatments are the least impressive part of the book. Of course there is little to say about insulin and E.C.T. over and above what one finds in almost any text-book nowadays, but even the psychological effects of cortisone are discussed in a turgid style, and one gets the impression that this part of the work is a piece of not very seasoned stuffing which has been thrown in for luck.

The final two sections on neurosurgery are very good. Here one finds in a very condensed form all the latest information on the effects of the various cutting operations for the relief of psychological symptoms.

OTHO FITZGERALD

Psychological Problems in Mental Deficiency. By SEYMOUR B. SARASON. New York: Harper. Pp. x+402. 40*s.*

This timely study of the difficulties of diagnosis of mental deficiency and its social and educational concomitants is as appropriate to the British scene as it is to the American. Work with mental defectives, although exacting, has often seemed unexciting and limited to ensuring that no physical or social disaster

occurs; this book shows that, for the defective, classification *per se* does not answer the real questions, that the limits of realization and development that can be achieved have not yet been reached and that psychotherapy may well be a profitable method of treatment.

There is included an analysis of the varieties of amentia and particular attention paid to the 'garden' subcultural variety—which approximates to simple feeblemindedness—for the environmental causation of which Dr Sarason makes out a good case, setting it in the wider framework of the relations between child-rearing practices and behavioural patterns, and illustrating his point by examples of the considerable facility by which individuals move in and out of classification as feebleminded.

Dr Sarason makes a careful and well-documented study of assessment procedures and test interpretation; he will know that he takes most people with him when he qualifies the value of a particular figure obtained on an intelligence test, but he also indicates the danger of using other tests such as the Rorschach, for this purpose, on the presumption that they yield all the relevant data for assessment purposes—the best test may reveal little of the potential of a subject or how he functions in other situations.

Dr Sarason sees the problem of the defective not merely as a social or educational nuisance but as one of adjustment that the individual concerned, his family and his social milieu have to make, and not the least important chapter is the one in which communication of information to parents is analysed and sample conversations set forth. Many misunderstandings, gaucheries and insensitive pronouncements will be avoided in the future if the problem is approached on Dr Sarason's realistic basis: 'The handling of the problems requires not only knowledge of certain facts but an understanding of what is essentially a psychotherapeutically-oriented study; a sustained attempt on the part of some trained person to understand the motives of parents, their frustration and hopes, and by virtue of such understanding as well as by previous training to enable parents to accept more realistic and satisfying attitudes.'

The least successful chapter of the book is that concerned with brain injury and aphasic conditions; for example, no reference is made to the work of Head, Adrian and Zangwill, but this is the only largely theoretical excursion in what is an essentially practical book. It can be confidently recommended to all clinical and educational psychologists and to psychiatrists, medical officers and social workers who are concerned with mental defectives and their parents.

JOHN M. GIBBS

Isaac and Oedipus. By ERICH WELLISCH. London: Routledge and Kegan Paul Ltd. 1954. Pp. xi+131. 15*s*. net.

This short study of the father–son relationship represents a re-evaluation of the Oedipus Complex in terms of the Biblical story of the Akedah (the unfulfilled Sacrifice of Isaac by his father Abraham). Wellisch's approach differs from Freud's in several respects. First, whereas Freud formulated the Oedipus Complex from the son's standpoint, Wellisch considers it in terms of the mutual relationship of father and son. Secondly, the primacy accorded by Freud to patricidal motives in the son is accorded by Wellisch to infanticidal motives in the father. Thirdly, whereas Freud's formulation was based on Greek mythology, Wellisch draws upon a Biblical story registering the historical abandonment of the ancient practice of infanticide in favour of a sacrificial rite in which the son can participate with the father. Fourthly, whereas Freud describes the resolution of the Oedipus conflict in terms of an identification of the ego with an introjected parental figure in the form of the super-ego, Wellisch regards this resolution as only partial and precarious unless there is included in the super-ego the introjection of an objective moral force promoting altruistic love ('the voice of God' which arrested the sacrifice of Isaac). Thus, in terms of the Akedah story, Wellisch envisages the possibility of 'salvation' from the Oedipus conflict in contrast to the stoical 'compromise solution' envisaged by Freud in conformity with the fatalism of the Greek tragedy of Oedipus. Whatever view the psychological reader may adopt towards the religious aspect of the thesis expounded in this book, he can hardly fail to be impressed by Wellisch's exposition of the psychological importance of infanticidal motives, supported as this is not only by rich anthropological evidence, but also by the internal evidence of the Oedipus myth itself (for Oedipus was exposed upon a mountain at the instance of his father's jealousy long before he killed his father). According to Freud patricidal, combined with incestuous, motives are the inevitable outcome of the instinctive endowment of the child; but the critical question raised by Wellisch's thesis is whether these motives are not largely engendered by the environmental influences of the family situation.

W. RONALD D. FAIRBAIRN

The Leaven of Love. By IZETTE DE FOREST. London: Gollancz. 1954. Pp. xv+206. 15*s*.

This is a simple and well-written book for the non-technical reader. It deals with 'The relation of the psychotherapeutic endeavour to the two Great Commandments, as exemplified in the theory and technique of Ferenczi'. 'Giving to the neurotic patient the hope and the healing experience of love makes way for a mutually loving relationship', the author points out on p. 19. 'This train of emotional circumstances in psychotherapeutic treatment has not as yet been adequately acknowledged or valued by the various schools of psychoanalysis. Yet for Sandor Ferenczi', she says, 'this was the sure measure of successful psychoanalytic therapy, the constant aim of his lifelong research.'

R. W. PICKFORD

Insanity, Art, and Culture. By FRANCIS REITMAN. Bristol: John Wright and Sons Ltd. 1954. Pp. 111. Price 12*s*. 6*d*.

This book may be viewed as a sequel to the same author's *Psychotic Art*, and embodies certain interesting researches which he has carried out on artistic work of mental patients, and a consideration of the relationships between their work and the cultural background. It has six illustrations and a frontispiece. The scope of the book may be gathered from a summary of the contents of its chapters. The first deals in a general way with the problems of psychotic art. The second deals with cultural influences, and after an examination of examples of psychotic art from a number of different cultural settings, suggests that such differences as are found in the art of psychotic patients in various cultures are due to cultural influences rather than to particular disease processes. The third chapter deals with the influence of the medium, namely that of embroidery or sculpture, used by patients, and it suggests that the use of the medium of sculpture tends to prevent the appearance of psychotic characteristics. The fourth chapter deals with an artist, George Morland, who was an alcoholic of manic-depressive temperament. It shows clearly that the art of such a man may remain unaffected by his psychotic tendencies. The fifth deals with psychic art, namely, automatic painting, doodles and spiritualistic products. Some features of these works are common to them and psychotic art, but such psychic art works are not obviously psychotic. They may be intermediate between psychotic and normal art. The sixth chapter deals with eccentricity and culture, and it discusses a well-known eccentric of the eighteenth century, John Mytton, a man of psychopathic personality, whose peculiarities were nevertheless socially acceptable in his time. In this chapter the author claims to show that unusual art may be acceptable culturally just like unusual behaviour. The last chapter is a summary.

The book suggests many interesting lines of research which might be followed out profitably concerning the relationships between culture, art and mental abnormalities. The author does not disguise the fact that all the problems dealt with in the book deserve much more detailed study. He has shown how material might be collected and handled and how statistical tests of significance might be applied to the differences of frequency observed. One feels that he has viewed a vast group of problems and applied various probes to the mass, and therefore his work is suggestive rather than conclusive, but it is unfortunate that he precluded himself from using the help which psychoanalysis might have given, by assuming that it is unscientific.

R. W. PICKFORD

Class Status and Power. A Reader in Social Stratification. Edited by REINHARD BENDIX and SEYMOUR MARTIN LIPSET. London: Routledge and Kegan Paul Ltd. 1954. Pp. 723. 42*s*. net.

Pages 17–45 of this volume are pure gold. They consist of reprints from classical writers, and it is delightful to learn from Aristotle that the value of the middle class resides in their ability both to command and to obey, to read Adam Smith pointing out the defects of capitalism in his beautiful, terse English, to discover that Marx believed people sometimes to act from motives other than love of gain, and to become acquainted with Veblen's surprising opinion about the value of good manners to their possessor. After that the book becomes verbose and uneven. Irritated by the lack of plan, the reader begins to wish either that Bendix and Lipset had written a book themselves, stating a few important hypotheses about social class, marshalling arguments for and against them, and quoting other writers sparingly, or, if they wished to make an anthology, that they had made a high quality of writing the test for inclusion.

One of the best contributions is Henry Pirenne's history of capitalism from the early Middle Ages to the present day, for, without hiding the complexity of the facts, he does seek to relate them in an orderly manner to a political and ethical thesis. Boris Brutzkus's outline of the economic history of Russia has a breadth of feeling that makes the reader sympathize with both the Tsarist régime and the revolutionaries, because he comes to understand the greatness of the difficulties with which they were battling. David Riesman propounds the interesting view that in present-day America neither generals nor trade union leaders behave as ruthlessly as their real power permits, because they are inhibited by inner psychological restraints, but he leaves his problem unresolved, because he has neither any clear opinion as to whether ruthlessness is good or bad, nor any awareness of the deeper motives underlying behaviour. In fact, practically nothing is anywhere explained in terms of the instinctual springs of action. There are, however, many thoughtful and informative essays, and a great quantity of statistics about various aspects of social class in the U.S.A.

JANE DARROCH

Motivation and Morale in Industry. By MORRIS S. VITELES. London: Staples. 1954. Pp. 510. Price 50*s*.

This must be the fruit of years of work. An enormous amount of sheer labour has gone into the task of compiling the results of hundreds of investigations; a remarkably clear mind has been at work collating it all and trying to turn it into a coherent whole. (Some of the chapters would make excellent models for the post-graduate student who is having difficulty with the section of his thesis headed 'Review of the Pertinent Literature'.) The smallest number of references quoted and listed in any chapter is 18, the largest 168. There are over 700 listed in the 'Index of Names' and 18 are British—which must be quite a record for American psychological works.

The book has five parts: I, Mobilizing the will to work; II, The background of motivational theory; III, Experimental studies; IV, Employee-attitude surveys; V, Guide posts for management. Part I is mainly concerned with setting and illustrating 'the problem'. Part II gives an admirably succinct account of relevant theories—incidentally putting Lewin and his associates and descendants clearly in the lead. Part III is a *tour de force* of compression. Part IV is not quite so good but uses many valuable sources other than psychological journals and is very informative. Part V, to my mind, makes a rather weak ending to the whole thing.

It was Part V which started me wondering whether I was reviewing *a book* or a work of reference. The preface states that it is meant primarily for management and only secondarily 'for courses in industrial management, personnel administration, industrial engineering and the like, as well as in courses in the field of psychology'. I think the primary and secondary purposes will be reversed in Great Britain. It is almost compulsory reading for all who talk about or practise industrial psychology; I cannot, however, see many British managers getting through the 500 pages—nor do I think they would get a great deal out of it all if they did. There is, of course, a theme to the book; it is printed on the spine. But (*a*) neither the problem-setting Part I nor the theoretical Part II is more than thinly related to the rest of the work, and (*b*) the conclusions of Part V do not follow with any logical force from the preceding pages. Viteles cannot be criticized for this: the fault lies in industrial psychological research—or, perhaps, in the awkwardness of human beings whose actions and reactions cannot at present be charted in any tidy way.

DENIS MCMAHON

First Report of the Joint Committee on Human Relations in Industry. (March 1953 to March 1954.) Department of Scientific and Industrial Research and Medical Research Council. London: H.M.S.O. 1954. Pp. 16. Price 1*s*.

Before the post-war Committee on Industrial Productivity was disbanded it recommended that the work of its Human Factors Panel should be taken over jointly by the Department of Scientific and Industrial Research and the Medical Research Council. Accordingly, in May 1953 two joint committees were appointed, one on Individual Efficiency in Industry, the other on Human Relations in Industry. The task of the Human Relations committee is to examine current researches in this field and to draw attention to problems on which research might be undertaken or supported by the Department of Scientific and Industrial Research, the Medical Research Council, or other bodies. At present the researches being wholly or partly supported fall into five main areas: Incentives, Management Organization, Promotion and Training, Special Groups in Industry (for example, married women), Technological

Change, and each of these areas is the special responsibility of a corresponding subcommittee. No details are yet available of the progress made by the researches.

As an example of how researches which are being sponsored interlock, the projects under the 'Incentives' heading include a critical evaluation of previous researches; a study of the impact of incentive schemes on productivity in terms of output, labour turnover, absenteeism, taking account also of trade union attitudes and experience; a study of the historical, economic and cultural factors affecting the establishment of production norms determined not so much by technical considerations as by attitudes of workers and management; a factory study of the process of changing from one method of payment to another; and finally an exploratory study of incentives for executive grades in industry.

As it happens, in this first year many of the researches are at least partly supported by Conditional Aid Programme funds; special conditions apply in that researches assisted by these funds must be directly related to productivity and industrial efficiency, and must show promise too of reaching applicable results before the end of the Conditional Aid Programme in 1956—conditions that do not decrease the difficulty of the Committee's task. However, although Conditional Aid has dominated the first year of the Committee's work and advanced it very considerably, the primary aim of the Committee is still to build up a 'balanced and continuing programme of social science research related to the needs of industry' as a bed-rock. Clearly one of its own main problems will be to identify the important topics for research and to assist in making the best use of the results. In identifying topics the perceived needs of industry form an important, though not by any means the only, criterion; the Committee's function as a link between industry and the existing social science research institutions is thus a vital one. A balance has yet to be found, however, between 'projects to aid the assimilation by industry of available knowledge, short-term applied research, and studies of a more fundamental character'.

R. M. MCKENZIE

The Modern Family. By ROBERT F. WINCH. New York: Henry Holt and Co. 1952. Pp. xxi+522. $3.90.

Winch's book is chiefly concerned with the modern American urban middle-class family; this is, however, set in perspective by describing other varieties of American families, both past and present, and even families belonging to non-American cultures, and by keeping in mind the changes that the chosen type of family is now undergoing. The first third of the book discusses the principal functions of the family; the second third, the relations between parent and infant, parent and child, parent and adolescent, and parent and adult, each of these getting a chapter to itself; the last third deals with courtship, love and marriage. This last section is probably the most original and interesting part of the book; it is certainly the most entertaining, especially the chapter on 'Romantic Love in American Life'. The author's aim was to discuss the family's influence on and share in the formation of personality. He has succeeded in maintaining a proper balance between the sociological and psychological aspects of his theme. The book is clearly written, well organized, and well documented, and can be recommended either as a survey for the general reader or as an introduction from which the student can readily pass to a closer study of special topics.

H. E. O. JAMES

From Loneliness to Fellowship. By WILHELM AAREK. London: Allen and Unwin. 1954. Pp. vi+79. 5*s*. net.

This little book is the published form of this year's Swarthmore Lecture, which is delivered at the beginning of each Yearly Meeting of the Society of Friends. The author is a Norwegian psychologist whose lecture is a thoughtful essay on man's need for fellowship and the means of achieving it. This has led him to discuss such topics as social integration, isolation, discipline and freedom, conscience, participation and so on, which he has done from a psychological point of view. But, oddly enough, it is the final chapter on 'Quaker Fellowship' which is most likely to arrest the attention of a psychologist, although this chapter ostensibly treats the problem from a religious standpoint. His account of the functioning of a Quaker Meeting, whether it be for worship or business, is worth some study by any psychologist who is interested in the dynamics of group life.

RALPH HETHERINGTON